

## A novel computer simulation method for simulating the multiscale transduction dynamics of signal proteins

Emanuel Peter, Bernhard Dick, and Stephan A. Baeurle

Citation: *J. Chem. Phys.* **136**, 124112 (2012); doi: 10.1063/1.3697370

View online: <http://dx.doi.org/10.1063/1.3697370>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i12>

Published by the [American Institute of Physics](#).

---

### Additional information on *J. Chem. Phys.*

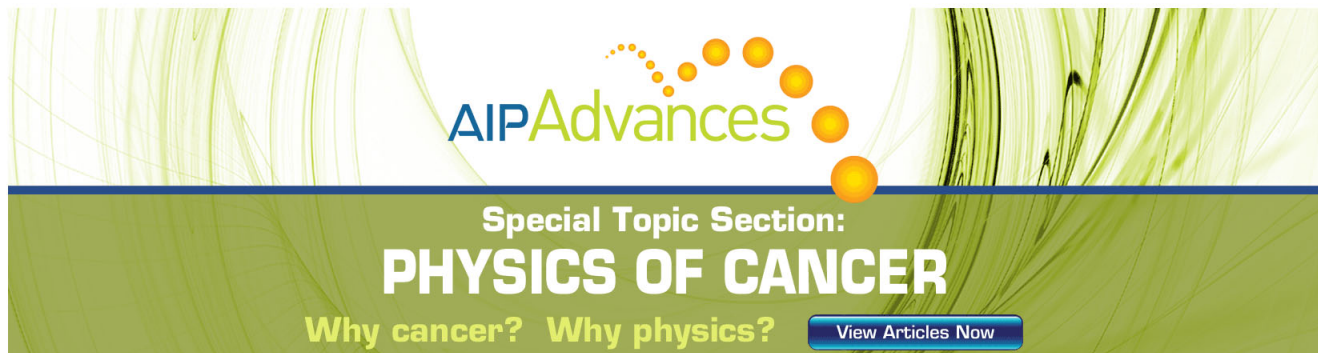
Journal Homepage: <http://jcp.aip.org/>

Journal Information: [http://jcp.aip.org/about/about\\_the\\_journal](http://jcp.aip.org/about/about_the_journal)

Top downloads: [http://jcp.aip.org/features/most\\_downloaded](http://jcp.aip.org/features/most_downloaded)

Information for Authors: <http://jcp.aip.org/authors>

## ADVERTISEMENT



**AIP Advances**

Special Topic Section:  
**PHYSICS OF CANCER**

Why cancer? Why physics? [View Articles Now](#)

# A novel computer simulation method for simulating the multiscale transduction dynamics of signal proteins

Emanuel Peter, Bernhard Dick, and Stephan A. Baeurle<sup>a)</sup>

Department of Chemistry and Pharmacy, Institute of Physical and Theoretical Chemistry, University of Regensburg, D-93040 Regensburg, Germany

(Received 21 November 2011; accepted 6 March 2012; published online 29 March 2012)

Signal proteins are able to adapt their response to a change in the environment, governing in this way a broad variety of important cellular processes in living systems. While conventional molecular-dynamics (MD) techniques can be used to explore the early signaling pathway of these protein systems at atomistic resolution, the high computational costs limit their usefulness for the elucidation of the multiscale transduction dynamics of most signaling processes, occurring on experimental timescales. To cope with the problem, we present in this paper a novel multiscale-modeling method, based on a combination of the kinetic Monte-Carlo- and MD-technique, and demonstrate its suitability for investigating the signaling behavior of the photoswitch light-oxygen-voltage-2- $J\alpha$  domain from *Avena Sativa* (AsLOV2- $J\alpha$ ) and an AsLOV2- $J\alpha$ -regulated photoactivable Rac1-GTPase (PA-Rac1), recently employed to control the motility of cancer cells through light stimulus. More specifically, we show that their signaling pathways begin with a residual re-arrangement and subsequent H-bond formation of amino acids near to the flavin-mononucleotide chromophore, causing a coupling between  $\beta$ -strands and subsequent detachment of a peripheral  $\alpha$ -helix from the AsLOV2-domain. In the case of the PA-Rac1 system we find that this latter process induces the release of the AsLOV2-inhibitor from the switchII-activation site of the GTPase, enabling signal activation through effector-protein binding. These applications demonstrate that our approach reliably reproduces the signaling pathways of complex signal proteins, ranging from nanoseconds up to seconds at affordable computational costs. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3697370>]

## I. INTRODUCTION

Signal-protein complexes act as regulators of the signaling pathways of cells in living systems, responding to a multitude of environmental stimuli such as light, temperature change, and/or mechanical stress.<sup>1</sup> A family of signal proteins of major importance in mammalian cells are the Rac-proteins, which are small GTPases that are involved in cell growth, cell-membrane adhesion, and cell survival.<sup>2</sup> Their activation is triggered by binding specific effector proteins at characteristic activation sites in the cell cycle. For example, the GTPase Rac1 is known to form aggregates with the Ser-Thr-protein kinase PAK1 at the activation site switchII, whose mutation is suspected to lead to deregulations inducing several malign tumors.<sup>3</sup> Signal proteins also play an important role in plants, in which they typically react in response to a change in illumination and/or temperature. For example, the phototropins are blue-light-sensitive protein complexes regulating a great diversity of biological processes, e.g., phototropic plant movement, chloroplast relocation, stomatal opening, rapid inhibition of stem growth, and gametogenesis, in higher plants as well as in micro-algae.<sup>4-7</sup> They are composed of two LOV domains, each containing a non-covalently bound flavin-mononucleotide (FMN) chromophore and a C-terminal Ser-Thr-kinase. Upon blue-light absorption, a covalent bond between the FMN chromophore and an adjacent reactive cysteine residue of the apo-protein is formed. This subsequently mediates the activation of the kinase, which induces a signal in the organism through phototropin auto-phosphorylation.<sup>8</sup> While the *in vivo* functionality of the LOV1 domain within the protein complex still remains unclear, the photochemical reactivity of the LOV2 domain has been found to be essential for the activation of the kinase.<sup>9</sup> As demonstrated in a series of experimental investigations,<sup>10-13</sup> the LOV2 domain releases upon illumination its inhibitory effect on the kinase by detaching a peripheral  $\alpha$ -helix, the so-called  $J\alpha$ -helix, from the LOV core. The early mechanism of activation was recently elucidated by us at atomistic resolution using MD simulation.<sup>14</sup> By connecting the AsLOV2- $J\alpha$  system and the previously mentioned Rac1-GTPase, Wu *et al.*<sup>15</sup> have lately created a genetically encoded fusion protein designated as photoactivable Rac1 (PA-Rac1), which makes use of the same photoswitching mechanism to modulate the signaling behavior of Rac1. In particular, they demonstrated that in the dark the AsLOV2 domain of PA-Rac1 inhibits Rac1 signaling by blocking its binding to the effector protein PAK1, whereas upon illumination this steric inhibition is released through detachment of the  $J\alpha$ -helix from the LOV core, enabling Rac1 activation. By further expressing PA-Rac1 in cancer cells of the HeLa-line, they could show that such protein constructs can be used to control the cell movement and functionality through light pulses.<sup>15</sup> Moreover, we note that the early mechanism of activation of this photoenzyme was recently elucidated by us at atomistic resolution using MD simulation.<sup>16</sup>

<sup>a)</sup>Electronic mail: stephan.baeurle@chemie.uni-regensburg.de.

To describe the structural-dynamics of such protein complexes on an atomistic length scale, several computational methodologies were developed starting from the late 1970s.<sup>17</sup> A prominent example among those is the MD technique, which describes the time-evolution of many-particle systems through phase space by numerically integrating Newton's equations of motion.<sup>18</sup> However, since its range of applicability for small proteins usually spans from nanoseconds up to sub-microseconds, its usefulness to study signaling processes on typical experimental timescales is only limited.<sup>19</sup> In order to reach longer timescales with the MD technique, several approaches have been proposed in the past decades. One of those is the coarse-graining (CG) approach, in which the system's degrees of freedom and, thus, the number of interactions is reduced enabling the use of larger time steps at lower computational costs. A successful CG method for protein systems is the united-atom approach, implemented, e.g., in the GROMOS96 forcefield.<sup>20</sup> It consists in representing all hydrogens with their respective aliphatic carbons as single effective atomic units. Approaches undertaking a more severe coarse-graining, such as the MARTINI method,<sup>21,22</sup> reliably reproduce protein structures but fail in providing the correct dynamics of complex protein systems, due to their strong heterogeneity at the atomistic level of description. To cope with the multiple length-scale problem of complex protein systems, several multiscale-modeling methods have been proposed, which partition a complex system in different space regions with varying degree of chemical resolution defined in an *ad hoc* fashion prior to the simulation. They rely on the idea of coupling theoretical methods with different levels of coarse-graining, i.e., quantum, atomistic, mesoscopic, or continuum-scale approaches,<sup>23–27</sup> within one simulation method. However, such techniques generally lack transferability, because they are specifically adapted to the nature of the physical problem under consideration and, thus, are not suitable for reproducing the multiscale relaxation dynamics of complex protein systems far from equilibrium. To extend the scope of computer simulation techniques to such non-equilibrium situations multiple-timestepping approaches, such as the reversible reference-system propagator algorithms (RESPAs)<sup>28</sup> or the Langevin dynamics integrator technique,<sup>29,30</sup> have been devised that greatly accelerate simulations of systems based on a separation of timescales and/or potential ranges. They generally rely on a decomposition of the dynamical range of the macromolecule into slow and fast modes, which allows an efficient propagation of the slower dynamical components through the use of larger time steps in the numerical integration procedure. With these approaches, an acceleration from 10 up to 100 times with regard to the conventional MD technique could be achieved, extending the scope of application of MD-based techniques to the microsecond or even sometimes to the sub-millisecond time range for peptides and small proteins.<sup>31,32</sup> Even if the gains in timescale are substantial, these techniques are still computationally very demanding, generally requiring several months of calculation on large parallel computer clusters<sup>32,33</sup> or supercomputers,<sup>31,34</sup> and therefore, will not be useful for the calculation of the signaling pathways of complex protein systems in the foreseeable future. Other techniques, such as the shadow hybrid

Monte Carlo method<sup>35</sup> or the meta-dynamics method,<sup>36,37</sup> enable a faster sampling of the free-energy surface of peptides and medium-sized proteins, using an effective Hamiltonian or modified potentials, respectively. These techniques permit to achieve a significant computational saving with regard to MD-based approaches, however, by construction are not able to reproduce the real dynamics of complex protein systems.

An alternative to the dynamical methods introduced previously are rate-based approaches, which mimic the time evolution of the system's trajectory through performing uncorrelated jumps from one state to another by circumventing activation barriers with a certain probability. One effective implementation of this conception is the kinetic Monte-Carlo (KMC) method, relying on the Bortz-Kalos-Lebowitz (BKL) algorithm.<sup>38</sup> This technique permits to generate a dynamical pathway, in which the most likely events are selected with higher probability from a list of events, whose rates have been previously determined from a simplified potential model. Several authors have applied this or related KMC techniques in the past few years to increase the accessible timescales in simulation studies of important protein folding<sup>39–41</sup> and unfolding problems.<sup>42,43</sup> For example, Makarov and co-workers<sup>42,43</sup> introduced a KMC-based algorithm to simulate the force-induced unfolding process of the muscle protein titin, using double-well potentials to describe the breakage and formation of H-bonds between different secondary-structure elements. In their algorithm the list of events is updated dynamically as the simulation proceeds, including all possible events associated with H-bond breakage or H-bond formation processes available at a particular instant of time during the simulation. From the list of events, the rates are selected using a modified BKL algorithm and the procedure is pursued until the protein is entirely unfolded. The rates are calculated through an Eyring-type formula, where the energy barriers are dependent on the pulling force acting on the H-bonds. With this approach, Makarov *et al.* were able to reproduce results from time-dependent pulling experiments with titin immunoglobulin domains, obtained by using single-molecule atomic force microscopy.<sup>44</sup> A major drawback of the KMC implementations discussed previously is that the rates are determined from simplified potential models, which do not reflect the full interaction spectrum of complex protein systems. In the present paper we introduce a novel algorithm, based on the combination of a BKL-type KMC- and a MD-algorithm. Our technique permits to calculate rates on the fly from an existing standard atomistic force field, optimized for protein simulations, and allows to extend the overall simulation time of conventional forcefield-based computer simulation techniques to typical timescales encountered in biological experiments.

Our paper is organized in the following way. In Sec. II, we present the methodology as well as the simulation parameters, followed by a description of the system preparation. Then, in Sec. III we discuss the simulation results on the AsLOV2-J $\alpha$ - and PA-Rac1-systems, in comparison to the available experimental and simulation data. Finally, we end our paper with a summary and a brief outlook.

## II. METHODS

To develop our simulation algorithm, we consider that the protein complex under consideration behaves as an infrequent-event system, in which the dynamics is characterized by occasional transitions from one state to another with long periods of relative inactivity between the transitions.<sup>45</sup> In this picture each state corresponds to a single energy basin and the long time between the transitions arises because the system must surmount an energy barrier to get from one state to another. The relaxation phase between the transitions ensures that the excess energy of the system can fully dissipate, allowing its thermal relaxation, and that the configuration becomes statistically uncorrelated to the previous one. It usually takes a few picoseconds or a few tens of picoseconds, i.e., several vibrational periods of a complex protein system, and can easily be carried out using a MD-type algorithm. The rates of all possible events out of each state  $i$  are calculated on the fly after each relaxation phase from a biomolecular forcefield and a list is created with them. The events are chosen according to the BKL algorithm,<sup>38</sup> which selects the events with a certain probability and rate corresponding to the most probable pathway. In our approach we assume that the multiscale transduction dynamics of complex protein systems is dominated by rate-determining steps, which according to Makarov and co-workers<sup>39,40,42</sup> are mainly associated with the breakage and formation of H-bonds in the protein system. We note that the stress concentration in these H-bonds is able to affect the energy barriers significantly, generally rendering the transitions more likely. Moreover, in our algorithm we assume that these events follow an Eyring-type mechanism, in which local fluctuations induce transitions of local events driving the subsequent global conformational change.<sup>46</sup> If all the previous requirements are fulfilled, the state-to-state dynamics of our protein system performs a Markov walk through state space, which means that the transition probabilities for exiting state  $i$  are independent of the history prior to entering state  $i$ . In this context it is also worth emphasizing that, if the rates are calculated exactly for each of the energy states visited, the state-to-state trajectory will be similar to the one calculated from a pure MD simulation at significantly lower computational costs.<sup>45</sup> A detailed description of our new algorithm and the resulting computational savings will be given in Secs. II A and II B.

### A. KMC-MD algorithm

To treat the different length- and time-scales of the complex protein system, we assume that the protein can be split into a network of dynamically heterogeneous entities,<sup>47</sup> each having their individual internal relaxation processes characterized by specific relaxation times. Moreover, we consider that, by going over to the high temperature limit, the protein unfolds and undergoes a transition from inhomogeneous to homogeneous kinetics by passing through the folding-unfolding transition temperature. This is caused by the fact that in this regime the landscape of the potential energy surface flattens and the relaxation behavior of all entities becomes dynamically homogeneous.<sup>48</sup> In the case of our LOV-

based protein complexes, we take into account that in the first stage the dynamics in the vicinity of the FMN chromophore, triggering the initial signal, is essentially dominated by the dynamics of single residues, which relax on very fast timescales in the range from picoseconds up to sub-nanoseconds. These residues have been identified in several experimental works<sup>49,50</sup> as being involved in the signaling pathway. The algorithm, describing this scale, will be designated in the following as the residue-based KMC (RB-KMC) procedure. In the second stage we consider the dynamical processes between different protein domains, each relaxing on timescales in the range from nanoseconds up to seconds. These can typically be secondary structure elements or large protein entities, which can easily be identified through a fluctuation analysis made from a short MD run prior to the KMC simulation. We call the KMC algorithm, treating this phase, the inter-domain KMC (ID-KMC) procedure. Both KMC algorithms generate a dynamical trajectory through state space by sampling the path of highest probability, guaranteed by the application of the BKL technique for the rate selection. To ensure that the trajectory satisfies the requirements of a Markov chain, the KMC walk from state to state is complemented by a MD-relaxation phase in each state, allowing the sampling of the new energy basin and the thermal relaxation of the system within a few picoseconds. We call the overall algorithm the KMC-MD approach, whose implementation and parameters will be discussed in more detail in Secs. II A and II B.

### 1. Residue-based KMC

This phase of the KMC algorithm treats the reorganization of the amino acids in the vicinity of the FMN chromophore, taking place in the timerange from picoseconds up to sub-nanoseconds. From several experimental and theoretical investigations,<sup>8,14,51-54</sup> it is well known that the three amino acids Gln513, Asn492, and Asn482 are essential in triggering the signaling pathway of the AsLOV2-J $\alpha$  as well as the PA-Rac1-system. We assume that these amino acids can carry out the following rate-dependent transformation processes, i.e., (1) side chain rotation, (2) H-bond formation, and (3) H-bond breakage, which are included in the list of events of the RB-KMC algorithm. Our approach relies on the BKL algorithm and is composed of the following steps:

1. MD phase;
2. H-bond analysis;
3. Scan of potential energy difference  $\Delta\Phi^{\ddagger}_{protein}$  along process path for each rate-determining event, which might take place in the simulation phase;
4. Calculation of rate of each event according to

$$r = \nu^{\ddagger} e^{\left(\frac{-\Delta G^{\ddagger}}{RT}\right)} \approx \bar{r}^{T \rightarrow \infty} e^{\left(\frac{-\Delta\Phi^{\ddagger}_{protein}}{RT}\right)}, \quad (1)$$

where  $\nu^{\ddagger}$  and  $R$  represent the pre-exponential factor and universal gas constant, respectively, whereas  $\bar{r}^{T \rightarrow \infty}$  designates the rate in the infinite temperature limit  $T \rightarrow \infty$ ;



5. Select event  $i$  from list of events by generating a uniform random number  $\xi \in (0, 1]$  and by solving

$$R_{i-1}/R_N < \xi \leq R_i/R_N, \quad (2)$$

where  $R_i = \sum_{j=1}^i r_j$  denotes a cumulative function with  $i = 1, \dots, N$  and  $N$  the total number of events;

6. Execute selected event  $i$  and update time through  $t = t + \Delta t$  by generating a uniform random number  $\xi \in (0, 1]$  and calculating

$$\Delta t = -\frac{\ln \xi}{R_N}; \quad (3)$$

7. Return to step (1).

In order to demonstrate that the approximation in Eq. (1) can be used to obtain a reliable estimate of the Gibbs energy of activation  $\Delta G^\ddagger$  in our algorithm, we assume that under isothermal conditions it adopts the following form:

$$\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger, \quad (4)$$

where  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  denote the enthalpy and entropy of activation of the process, respectively. This permits to reformulate the rate expression in Eq. (1) as<sup>55</sup>

$$r = \bar{r}(T)e^{\left(-\frac{\Delta H^\ddagger}{RT}\right)} \quad (5)$$

with

$$\bar{r}(T) = v^\ddagger e^{\left(\frac{\Delta S^\ddagger}{R}\right)}. \quad (6)$$

In Eq. (5), the temperature dependence of the pre-exponential term is generally negligible in comparison to the much stronger temperature dependence of the exponential term.<sup>56</sup> Using this assumption, we can re-write the rate as follows:

$$r \approx \bar{r}^T \rightarrow \infty e^{\left(-\frac{\Delta H^\ddagger}{RT}\right)} \quad (7)$$

with the enthalpy of activation given by<sup>57</sup>

$$\Delta H^\ddagger = \Delta U^\ddagger + p\Delta V^\ddagger + \Delta pV^\ddagger. \quad (8)$$

Here, we note that  $\Delta p$  and  $\Delta V^\ddagger$  represent, respectively, the changes of the external pressure and volume of activation, whereas  $\Delta U^\ddagger$  is the change of the internal energy of activation defined as

$$\Delta U^\ddagger = \Delta K^\ddagger + \Delta \Phi^\ddagger, \quad (9)$$

where  $\Delta K^\ddagger$  and  $\Delta \Phi^\ddagger$  designate the changes in the kinetic and potential energy of activation, respectively. The change in the potential energy  $\Delta \Phi^\ddagger$ , associated with each KMC event, can further be decomposed in the following contributions:

$$\Delta \Phi^\ddagger = \Delta \Phi^{\ddagger/protein} + \Delta \Phi^{\ddagger/solvent} + \Delta \Phi^{\ddagger/interface}, \quad (10)$$

where  $\Delta \Phi^{\ddagger/protein}$ ,  $\Delta \Phi^{\ddagger/solvent}$ , and  $\Delta \Phi^{\ddagger/interface}$  designate the changes in the potential of the protein, solvent, and protein-solvent interface of activation, respectively. To obtain a reliable approximation of  $\Delta H^\ddagger$ , we executed each event under solvent-free conditions at zero temperature. Then, we re-inserted the solvent and equilibrated it in a short canonical MD phase by freezing the degrees of freedom of the protein and adjusting the temperature related average kinetic energy

of the solvent to its external value. This implementation ensured that for each executed event  $\Delta K^\ddagger \approx 0$ ,  $\Delta p \approx 0$ , and  $\Delta V^\ddagger \approx 0$ , leading to

$$\Delta H^\ddagger \approx \Delta \Phi^\ddagger. \quad (11)$$

Moreover, by considering that in protein-solvent systems the potential energy contributions related to the bonded and non-bonded interactions of the protein during the event execution dominate in magnitude with regard to the potential energy changes of the solvent and protein-solvent-interface, we can finally approximate the enthalpy of activation as

$$\Delta H^\ddagger \approx \Delta \Phi^{\ddagger/protein}. \quad (12)$$

By computing the signaling behavior of the AsLOV2-J $\alpha$ - and PA-Rac1-systems together with their mutants and comparing the results to experiments, we will show in the following that this implementation permits to reliably approximate  $\Delta H^\ddagger$  as the finite difference between the absolute maximum, i.e., the so-called transition state, and the absolute minimum prior to the transition state of the one-dimensional potential energy profile. To evaluate the profiles of all possible events during a KMC cycle, we accomplished the scan stepwise within a scanning range  $d_{scan}$  and completed each step by subsequent minimization using the Broyden-Fletcher-Goldfarb-Shannon algorithm,<sup>58</sup> while constraining the donor-acceptor distance of the amino acids under consideration. The hopping range for the execution of each event was determined in an adaptive way as the absolute minimum in the potential energy after the transition state and, subsequently, the selected event was executed. Then, the algorithmic cycle was re-started with a MD phase within the *NPT* ensemble by imposing periodic boundary conditions.

## 2. Inter-domain KMC

The second KMC phase considers rate-dependent processes between different protein domains, occurring on timescales in the range from nanoseconds up to several seconds. These domains can either be secondary structure elements or larger protein entities, which can be identified as dynamically homogeneous regions of the protein by means of a fluctuational analysis prior to the KMC-MD simulation. The ID-KMC phase is accomplished by carrying out the same algorithmic steps as in the RB-KMC algorithm, introduced in Sec. II A 1. In the case of the AsLOV2-J $\alpha$  system we consider the LOV2 core and the J $\alpha$ -helix as the dynamically homogeneous entities, supplemented in the case of the PA-Rac1 system with the LOV2-GTPase-interfacial region. Furthermore, we assume that the corresponding inter-domain relaxation is dominated by the rate-dependent processes of H-bond breakage. This latter simplification takes into account that the H-bond formation and side chain rotation processes are much faster compared to the H-bond breakage processes and, therefore, the former processes can be neglected on the larger timescales, addressed in the ID-KMC phase. This allowed us to reduce the number of KMC cycles and, thus, the costs of the overall calculation significantly.

## B. Simulation details and system preparation

### 1. Simulation details

We performed five independent simulations for each state of the AsLOV2-J $\alpha$ - and PA-Rac1-system to check the reproducibility of the simulations. The length of the MD phase between KMC steps of the KMC-MD simulations was determined through measurement of the statistical inefficiency of the MD-simulation algorithm in the case of the AsLOV2-J $\alpha$ - as well as PA-Rac1-system. This latter parameter represents the length of the MD trajectory, which is needed to get statistical independent configurations required by the KMC procedure. To compute this quantity, we used the method of Fincham *et al.*<sup>59</sup> and evaluated a length of the MD phase of 5 ps. Another important parameter, which needs to be determined prior to the calculation, is the maximum scanning range  $d_{scan}$  for the processes of H-bond formation and H-bond breakage as well as the maximum angle of rotation  $\alpha_{scan}$  for the side chain rotation processes. To this end, we performed independent MD simulations on the AsLOV2-J $\alpha$  system and evaluated for both H-bonding processes an optimal value of 0.35 nm from the average translational fluctuations, whereas for the process of side chain rotation we estimated a value of 20°. As illustrated in the Fig. 1SA of the supplementary

material,<sup>60</sup> we determined the process path of each possible H-bond breakage and H-bond formation event  $i$ , which could take place in the protein during KMC-MD run, through the following equation:

$$\mathbf{x}_i^{end/donor} = \mathbf{x}_i^{start/donor} + d_{scan} \frac{(\mathbf{x}_i^{start/donor} - \mathbf{x}_i^{acceptor})}{|\mathbf{x}_i^{start/donor} - \mathbf{x}_i^{acceptor}|}, \quad (13)$$

where  $\mathbf{x}_i^{start/donor}$  and  $\mathbf{x}_i^{end/donor}$  denote the starting and final coordinates of the donor, whereas  $\mathbf{x}_i^{acceptor}$  is the coordinate of the acceptor kept at a fixed position in space during the execution of the event. In Fig. 1SB, we show corresponding representative potential energy profiles for a H-bond breakage process in case of the dark and light states of the AsLOV2-J $\alpha$ -system. Moreover, the process path of the side chain rotation in three-dimensional space was defined through the following operation of rotation by an angle  $\alpha$  about an axis in the direction of the unit vector  $\mathbf{n} = (n_1, n_2, n_3)^T$ :

$$R_{\mathbf{n}}(\alpha)\mathbf{x} = \mathbf{n}(\mathbf{n} \cdot \mathbf{x}) + \cos \alpha(\mathbf{n} \times \mathbf{x}) \times \mathbf{n} + \sin \alpha(\mathbf{n} \times \mathbf{x}), \quad (14)$$

where the rotation matrix is given by

$$R_{\mathbf{n}}(\alpha) = \begin{pmatrix} \cos \alpha + n_1^2(1 - \cos \alpha) & n_1 n_2(1 - \cos \alpha) - n_3 \sin \alpha & n_1 n_3(1 - \cos \alpha) + n_2 \sin \alpha \\ n_2 n_1(1 - \cos \alpha) + n_3 \sin \alpha & \cos \alpha + n_2^2(1 - \cos \alpha) & n_2 n_3(1 - \cos \alpha) - n_1 \sin \alpha \\ n_3 n_1(1 - \cos \alpha) - n_2 \sin \alpha & n_3 n_2(1 - \cos \alpha) + n_1 \sin \alpha & \cos \alpha + n_3^2(1 - \cos \alpha) \end{pmatrix}. \quad (15)$$

More specifically, in our algorithm we chose in case of the amino acids Gln513, Asn492, and Asn482 in vicinity to the FMN the axis of rotation in the direction of the CA-CB axis (see Fig. 1SC) and performed a stepwise rotation up to the maximum angle of rotation  $\alpha_{scan}$ . We note that each execution of an event along a process path was complemented by a subsequent MD-relaxation phase, which reduces the dependence of our algorithm on the process path realization. As we will demonstrate in the following, this implementation was found to provide reliable results for the AsLOV2-J $\alpha$ - and PA-Rac1-systems under consideration and to be the most effective procedure with regard to computational expense. Moreover, we accomplished H-bond analysis, using standard geometrical criteria obtained from Refs. 61 to 63, i.e., a maximum donor-acceptor distance of 0.35 nm and a maximum hydrogen-donor-acceptor angle of 30°. For the determination of  $\Phi^{protein}$  during the KMC phase, we used a shift function with a cutoff of 1.4 nm for calculating the van-der-Waals (vdW) and electrostatic-potential energies. To determine the pre-exponential factor  $\bar{r}^{T \rightarrow \infty}$  in Eq. (1), we considered that, as described in Sec. II A, the protein unfolds and undergoes a transition from inhomogeneous to homogeneous kinetics by passing through the folding-unfolding transition

temperature with increasing temperature, which causes that in the infinite temperature limit all events possess identical rates designated by  $\bar{r}^{T \rightarrow \infty}$ . We point out that the latter assumption has been confirmed by several theoretical and experimental works, which will be discussed in the following. For example, Leite *et al.*<sup>48</sup> have studied fluctuations of protein folding kinetics using a lattice-based dynamic Monte Carlo algorithm by examining the ratios of the moments of the first-passage time. They found that at high temperatures, due to the large thermal fluctuations of the dynamical entities, local details are smeared out and the different kinetic paths sense roughly similar barriers, resulting in single-exponential kinetics. Moreover, Zhang *et al.*<sup>64</sup> observed by performing neutron scattering spectroscopy experiments and MD simulations on the hen egg white protein lysozyme that in the high temperature regime the lifetime of the H-bonds between the protein and the hydration water decreases substantially, which causes an increased mobility of the protein in the water environment and leads to its subsequent unfolding. Ultimately, in the high temperature limit the protein loses the remaining dynamical heterogeneity related to its interaction with the hydration water and adopts a homogeneous kinetic behavior, which is reflected by an Arrhenius-type temperature dependence of the

inverse diffusion constant.<sup>64</sup> Finally, we emphasize that the assumption of homogeneous kinetics in the high-temperature limit has also been successfully employed by Makarov and co-workers,<sup>39,40,42</sup> to reproduce pulling experiments on the muscle protein titin,<sup>44</sup> using a KMC method based on a modified BKL algorithm. Despite the ineffective infrequent-event sampling of the pure MD technique compared to the KMC-MD technique, we consider that the high local stress concentration in the binding pocket resulting from the formation of the cysteinyl-FMN adduct allows to reduce the activation barriers for the events, involving the residues near to the FMN chromophore in the AsLOV2 core. This ensures that the signaling pathway of both simulation methods can be assumed to be similar in the early stages. As a consequence, we can estimate the pre-exponential factor  $\bar{r}^T \rightarrow \infty$  by identifying similar characteristic events from the inter-atomic distances of the amino acids adjacent to the FMN chromophore, calculated with the KMC-MD and MD techniques, and adjusting the KMC-MD- onto the MD-time. To demonstrate this, we consider in Figs. 2SA and 2SB the inter-atomic distances between Gln513-OE1 and Asn492-ND2 of the PA-Rac1 system, obtained with the MD as well as KMC-MD techniques, respectively. We compare these results with the corresponding ones for the distance between Gln513-NE2 and Asn492-OD1, visualized in Figs. 2SC and 2SD. From the graphs, we deduce that at the simulation stage, marked with a black arrow, a permanent coupling through H-bond formation between Asn492 and Gln513 does occur, which triggers the N-terminal disruption of the  $J\alpha$ -helix of the PA-Rac1 system. The latter finding is confirmed by comparing the secondary structure analysis of the  $J\alpha$ -helix of the PA-Rac1 system, obtained with both simulation techniques and shown in Figs. 2SE and 2SF. By considering that the early signaling pathways of both simulation techniques are similar, we conclude that these events must be identical and that, therefore, the fifth KMC step must correspond to a real time of about 20 ns. Using this procedure, we obtained in case of the PA-Rac1 system a value for the pre-exponential factor of  $3.94 \times 10^7 \text{ s}^{-1}$ . In case of the AsLOV2- $J\alpha$  system, we used the same pre-exponential factor as for the PA-Rac1 system by taking into account that both signaling pathways are triggered through cysteinyl-FMN-adduct formation in the AsLOV2 core and, thus, they should be similar in the early stages. To carry out the MD-simulation phase at a molecular level, we used the GROMACS molecular dynamics simulation package Version 3.3.1 in conjunction with the GROMOS96-43a1 forcefield<sup>65</sup> to describe the interactions. This widely used forcefield has been tested against NMR-spectroscopic data in case of the lysozyme protein in water by Soares *et al.*<sup>66</sup> and has been found to reproduce its solution structure and conformational behavior very well. In a recent work, Todorova *et al.*<sup>67</sup> performed extensive MD simulations on the 51-amino-acid protein insulin and subjected the GROMOS96-43A1 forcefield to a systematic comparison against other popular biomolecular forcefields, including the CHARMM27-, AMBER03-, OPLS-, and GROMOS96-53A6-forcefields. They analyzed in detail the effect of each forcefield on the conformational evolution and structural properties of the protein and compared the results with the available experimental data. They observed

that each forcefield favors different structural trends. Moreover, they found that the united-atom forcefield GROMOS96-43A1, together with the CHARMM27-forcefield, delivered the best description of the experimentally observed dynamic behavior of the chain B of insulin. In our simulations we used in addition full particle-mesh-Ewald electrostatics with a Coulomb cutoff of 1.4 nm and computed the vdW interactions, using a shift function with a cutoff of 1.4 nm. In case of the AsLOV2- $J\alpha$  system we centered the protein into a cubic box of 6.47 nm and filled it with SPC-water as well as 4 sodium ions to electrically neutralize the system, whereas in case of the PA-Rac1 system we proceeded in the same way using a boxlength of 9.7 nm and 7 sodium ions. To generate an isothermal-isobaric ensemble with a temperature of 300 K and a pressure of 1 atm, we used a Nosé-Hoover thermostat and Parrinello-Rahman barostat.<sup>68</sup> To mimic non-equilibrium conditions, we decoupled the FMN chromophore and the protein from their individual thermostats, whereas the solvent and ions remained coupled to their respective thermostat throughout the whole simulation run. This technique is known as the non-invasive thermostating technique, which allows the protein to sample configurations far from equilibrium and to follow its natural dynamics under solvent-mediated thermostating control.<sup>53</sup> For the numerical integration of the equations of motion, we used the leapfrog integrator with a time step of 1 fs. To describe the interactions of the cysteinyl-FMN adduct entity, we used the parameters of Neiss and Saalfrank,<sup>69</sup> which were determined from B3LYP-6-31G\*-calculation results and through comparison with similar groups in the forcefield to reach consistency. The reliability of the parametrization of the cysteinyl-FMN entity was tested and confirmed by these authors on the LOV2 domain from *Adiantum capillus-veneris* without the  $J\alpha$ -helix<sup>69</sup> and by us on the AsLOV2- $J\alpha$  system<sup>14</sup> as well as the LOV1 domain of *Chlamydomonas reinhardtii*.<sup>52,53</sup>

## 2. System preparation

As starting structures for the dark-state simulations, we used the dark-state crystal structure of the PA-Rac1- (PDB-code: 2WKP) (Ref. 15) and AsLOV2- $J\alpha$  system (PDB-code: 2V0U),<sup>10</sup> visualized in Figs. 1(a)–1(d). To create the initial structures of our adduct-state simulations, we employed the dark-state structures and generated the cysteinyl-FMN (CFN) adduct by forming a covalent bond between Cys450-S and FMN-C4a. We point out that the covalent linkage between the reactive cysteine and FMN is required to transmit the stress from the reaction center to the protein and trigger the protein signal, as we will demonstrate in the following through comparison between dark- and light-state simulations and as we have shown in previous simulation studies with various LOV domains.<sup>14,16,52,54</sup>

## III. RESULTS AND DISCUSSION

To demonstrate the reliability and potential of our KMC-MD approach, we start by considering the structural-dynamical changes of the AsLOV2- $J\alpha$  system in the dark and



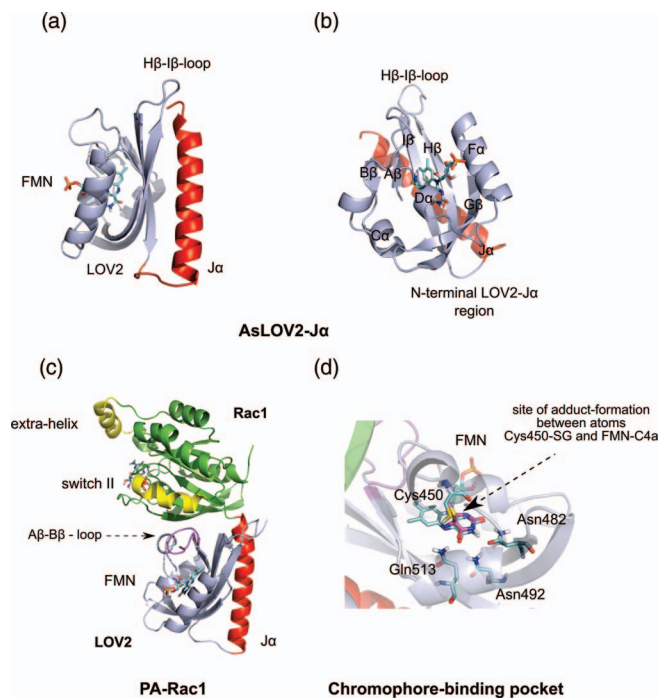


FIG. 1. Crystal structures of AsLOV2-J $\alpha$  and PA-Rac1 system (PDB-codes: 2V0U, 2WKP), determined through x-ray diffraction measurements by Halavaty and Moffat and Wu *et al.*, respectively. These are used as input structures for KMC-MD- as well as MD-simulations. (a) Side-view of AsLOV2-J $\alpha$  system with FMN chromophore. (b) Front-view of AsLOV2-J $\alpha$  system with different secondary-structure elements. (c) PA-Rac1 with activation site switchII on Rac1 enzyme connected with LOV2 core through J $\alpha$ -helix. (d) Amino-acid environment of FMN chromophore in AsLOV2 core [Gln513 (I $\beta$ ), Asn492 (H $\beta$ ), Asn482 (G $\beta$ )].

light state on the nanosecond timescale. In Fig. 2, we display the results for the amino-acid configuration around the FMN chromophore and its consequences on the peripheral J $\alpha$ -helix, obtained with our KMC-MD method, and compare them to the results, obtained from a 20 ns MD-simulation run with the same system. To begin, we show in Fig. 2(a) the distances between Gln513-NE2 and Asn492-OD1 for the dark and light states in the RB-KMC-MD phase. In case of the light-state curve, we see a clear drop in the distance at the 26th KMC step indicating H-bond formation, whereas the dark-state curve remains at a constant distance of 0.55 nm up to the end of this phase. To further illustrate the coupling between Gln513 and Asn492 in the KMC simulation, we show a representative configuration after 40 RB-KMC steps in Fig. 2(b). We observe a clear H-bonding between Gln513-NE2 and Asn492-OD1, which is a result of the rotation of the side chain of Gln513. This finding is confirmed by our recent theoretical investigation in Ref. 14 and various experimental works,<sup>8,13,70,71</sup> which provided evidence for the important role of Gln513 and its coupling with Asn492 in triggering the signaling pathway of the AsLOV2-J $\alpha$  domain. In particular, Nash *et al.* prove through mutational analysis that the Gln513 switch is directly implicated in the subsequent detachment of the J $\alpha$ -helix from the LOV core,<sup>70</sup> which is supported by our recent MD simulations on the AsLOV2-J $\alpha$  system.<sup>14</sup> By contrast, we see in the corresponding MD-simulation results of the light state of the AsLOV2-J $\alpha$  system in Fig. 2(c) that

no coupling occurs between Gln513-NE2 and Asn492-OD1. This is due to the limited MD time of 20 ns, in which the Gln513 switch and H-bond formation does not take place. In Fig. 2(d), we display a representative configuration of the chromophore environment after 20 ns of MD simulation, which shows that the amino acids Gln513 and Asn492 remain disconnected. We explain the discrepancy between the KMC-MD- and MD-simulation results by the difference in timescales accessed by both simulation approaches, which amounts to 130 ns at the end of the RB-KMC-MD phase. We point out that, due to the higher computational expense per time step, the MD-simulation approach significantly depends on the quality of the starting structure, as can be deduced by comparing the current MD-simulation results based on a crystal starting structure with the ones from our previous MD-simulation work, where we used a refined NMR-solution structure.<sup>14,72</sup> Next, we display in Fig. 2(e) the consequence of the H-bond coupling between Gln513 and Asn492 onto the tertiary structure of the AsLOV2-J $\alpha$  system. By comparing the structures of the dark- and light-states from the last KMC step of the RB-KMC-MD phase, we see that the N-terminal region of the J $\alpha$ -helix in the light state (bright red) is disrupted and bent, whereas in the dark state (dark red) it is similarly arranged as in case of the starting structure. The disruption of the J $\alpha$ -helix is caused by the coupling between the I $\beta$ - and H $\beta$ -strands, mediated through the H-bond formation between the residues Gln513 and Asn492, as shown in Figs. 2(a) and 2(b). By contrast, we note that the tertiary structure of the light state after 20 ns of MD simulation in Fig. 2(f) resembles the starting crystal or average dark state structures. From this fact, we deduce that a suitable coupling between Asn492 and Gln513 is a crucial step for the N-terminal disruption of the J $\alpha$ -helix, because no coupling occurred in case of the MD simulation approach, as demonstrated in Figs. 2(c) and 2(d). This is confirmed by Nash *et al.*,<sup>70</sup> who performed CD- and HSQC-NMR-spectroscopy measurements on the Gln513Asn-AsLOV2 mutant. Through their experiments they demonstrated, that this mutant does not exhibit light-state signaling behavior, which might be explained by the smaller side chain of Asn513 in comparison to Gln513. To study this aspect, we performed two independent KMC-MD simulations with the Gln513Asn-AsLOV2 mutant and found that already after a few simulation steps H-bond coupling occurs between Asn513 and Asn492. However, we observed that this coupling does not lead to a disruption of the J $\alpha$ -helix from the LOV2 core due to the shorter side chain of Asn513, as shown in Fig. 3, which is in agreement with the experiments of Nash *et al.* Next, we continue with the analysis of the results from the subsequent ID-KMC-MD phase, which follows the RB-KMC-MD phase. In Fig. 4(a), we show the final structures from two independent KMC-simulation runs of the dark state, in which total simulation times of 149  $\mu$ s and 624 s have been reached, respectively. Both structures have been overlaid using the smallest root-mean-square deviation between the atoms. From this figure, we deduce that the J $\alpha$ -helix remains attached to the LOV core until the end of the ID-KMC-MD phase. In contrast to that, we see in Fig. 4(b) that the J $\alpha$ -helix in the structures of the two light-state simulations is disrupted, which correspond to final simulation



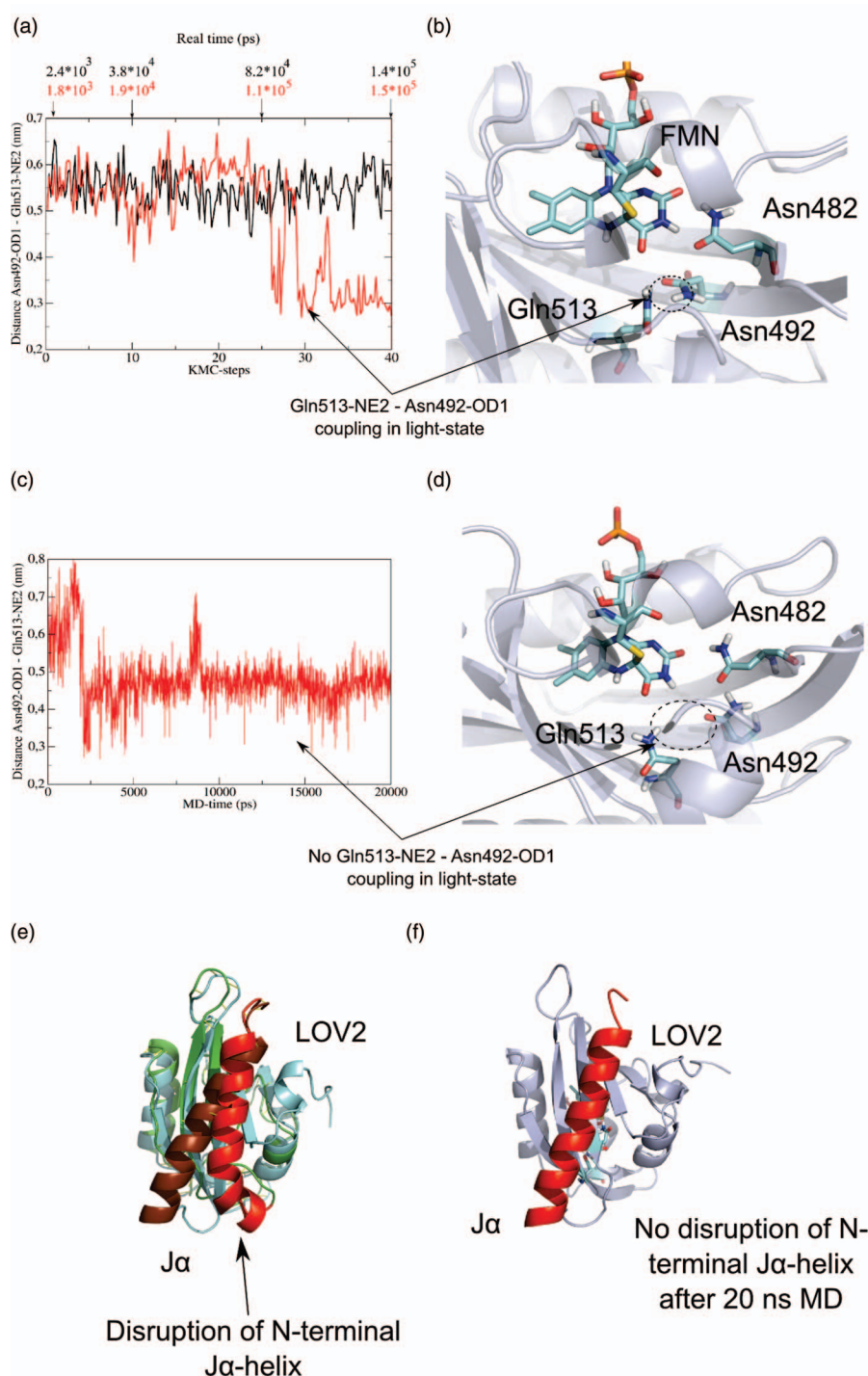


FIG. 2. Inter-atomic distances and representative structures of the AsLOV2-J $\alpha$  system, obtained by using KMC-MD- and non-invasive MD-methods. (a) Inter-atomic distance between Gln513-NE2 and Asn492-OD1 from KMC-MD simulation in the RB-KMC phase of the light state (red curve) and dark state (black curve) as a function of KMC steps. (b) Amino-acid configuration around FMN chromophore of the light state from KMC-MD simulation after 40 RB-KMC steps. (c) Inter-atomic distance Gln513-NE2 and Asn492-OD1 of the light state from MD simulation as a function of MD time. (d) Amino-acid configuration around FMN chromophore of the light state after 20 ns of MD simulation. (e) Overlaid structures of dark state (dark red) and light state (bright red) of AsLOV2-J $\alpha$  system after 40 RB-KMC steps. (f) Final structure after 20 ns of MD simulation.

times of 330  $\mu$ s and 127 s, respectively. From these observations, we conclude that in the light state the signal propagates from the amino acids near to the FMN chromophore to the peripheral AsLOV2-J $\alpha$ -interface region leading to the detachment of the J $\alpha$ -helix from the LOV core, whereas in the inactive dark state the J $\alpha$ -helix remains attached to the LOV core. This behavior is confirmed by several experi-

mental works, using x-ray-, 2D-HSQC-NMR-, and FTIR-techniques.<sup>8,10-12,73</sup> For example, Harper *et al.* demonstrated through 2D-HSQC-NMR experiments that CFN-adduct formation induces the disruption of the J $\alpha$ -helix from the LOV2 core.<sup>11</sup> Nakasone *et al.* performed time-resolved transient-grating- and transient-lens-experiments to determine diffusive changes of the LOV2-J $\alpha$  system from *Arabidopsis* upon light

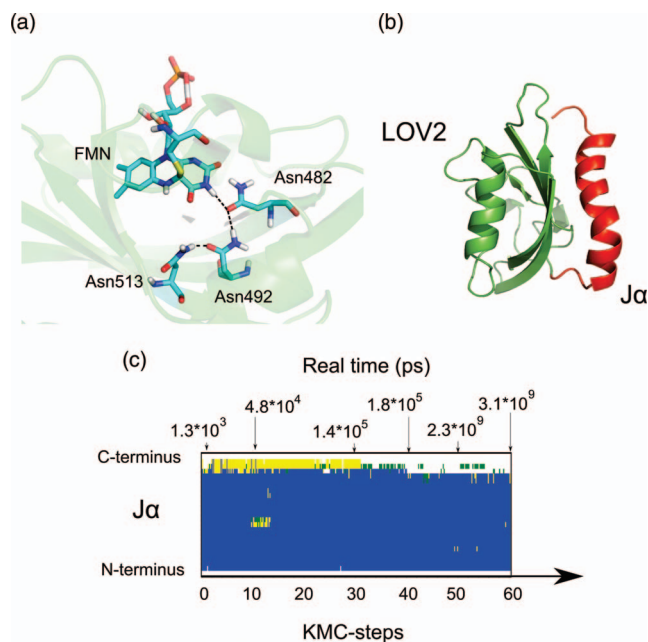


FIG. 3. (a) Final amino-acid configuration in vicinity of the FMN chromophore, (b) final overall structure as well as (c) secondary-structure analysis of the  $J\alpha$ -helix as a function of KMC steps and real time from Gln513Asn-AsLOV2 mutant in the light state, obtained with KMC-MD method.

excitation.<sup>19</sup> In their paper, they speculated that complete unfolding of the  $J\alpha$ -helix takes place after cleavage from the LOV2 core. To investigate the detachment and subsequent unfolding process of the  $J\alpha$ -helix in more detail, we visualize in Figs. 4(c) and 4(d) the secondary-structure analysis of the  $J\alpha$ -helix from the AsLOV2- $J\alpha$  system in the dark and light state over the entire simulation run, obtained with the KMC-MD method using the crystal structure of Halavaty and Moffat,<sup>10</sup> as starting structure. We compare these results with the ones obtained with the MD simulation technique, using the refined NMR-solution structure of Harper *et al.*<sup>11</sup> in Figs. 4(e) and 4(f) as well as with the crystal structure of Halavaty and Moffat<sup>10</sup> in Fig. 4(h). While in the dark state the  $J\alpha$ -helix remains folded with both simulation strategies, we see that the length of the folded sequence of the  $J\alpha$ -helix decreases in case of the light-state KMC-MD simulation as well as the MD simulation with the NMR-solution structure, whereas in case of the MD simulation with the crystal starting structure this behavior is not observed. By further comparing these results with the secondary structure analysis from the MD simulation run of the isolated  $J\alpha$ -helix in Fig. 4(g), we infer that after the coupling of the H $\beta$ - and I $\beta$ -strands the  $J\alpha$ -helix partially unfolds in the N-terminal region and after its complete cleavage from the LOV core the unfolding continues to a minor extent at the C-terminus. We explain the partial unfolding of the  $J\alpha$ -helix by the fact that no force is exerted by the C-terminal region of the LOV core, which could completely unfold the  $J\alpha$ -helix after full cleavage. This causes that the  $J\alpha$ -helix of the AsLOV2- $J\alpha$  system in the disrupted signaling state behaves similarly as the isolated  $J\alpha$ -helix in solution. This finding is confirmed by the FTIR measurements of Alexandre *et al.*,<sup>71</sup> who demonstrated that the  $J\alpha$ -helix unfolds only partially upon illumination. We point out that the partial unfold-

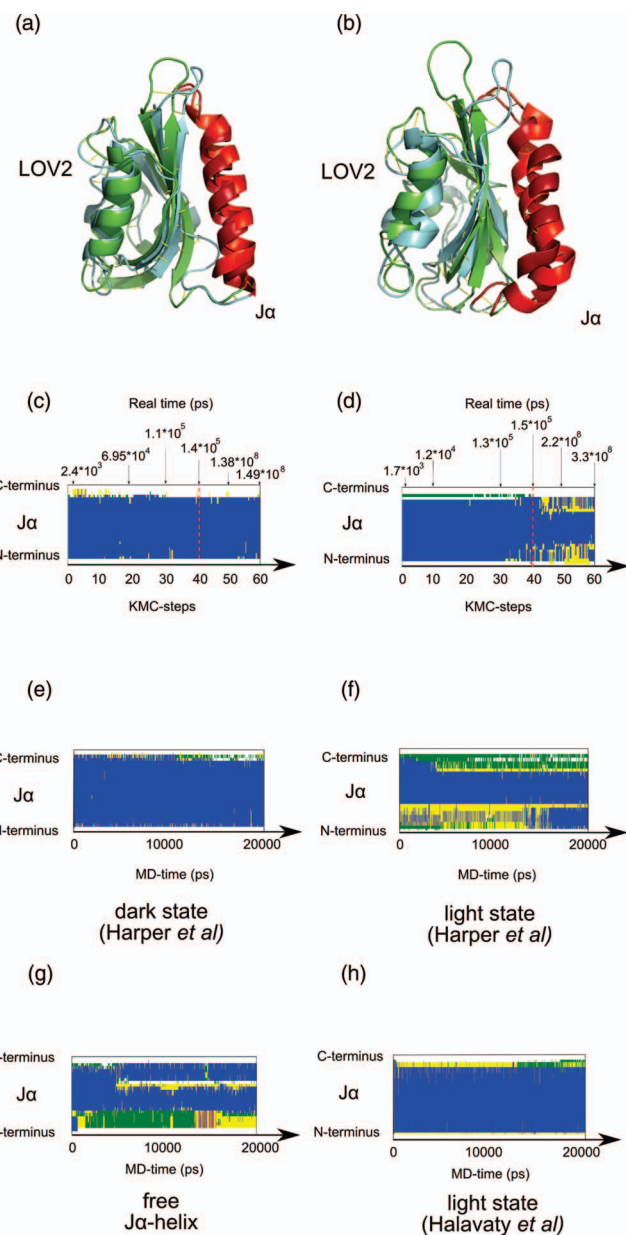


FIG. 4. Final structures of the AsLOV2- $J\alpha$  system in the dark state (a) and light state (b) after the ID-KMC phase at the end of KMC-MD simulation. Secondary-structure analysis of the  $J\alpha$ -helix of the AsLOV2- $J\alpha$  system in the dark state (c) and light state (d) as a function of KMC steps and real time, obtained by using KMC-MD method. Secondary-structure analysis of the  $J\alpha$ -helix of the AsLOV2- $J\alpha$  system in the dark state (e) and light state (f) as a function of MD steps, obtained by using MD method in conjunction with the NMR-solution structure of Harper *et al.*<sup>12</sup> as initial structure. Secondary-structure analysis of the  $J\alpha$ -helix of the AsLOV2- $J\alpha$  system in the free state (g) and light state (h) as a function of MD steps, obtained by using MD method in conjunction with the x-ray crystal structure of Halavaty and Moffat<sup>10</sup> as initial structure.

ing at the C-terminal end of the  $J\alpha$ -helix might be a consequence of the high mobility of the  $J\alpha$ -helix at this free chain end, caused by the missing kinase enzyme. This is confirmed by the fact that this behavior is not observed in the PA-Rac1 system discussed in the following, where such kinetic effects are excluded. The importance of the force at the C-terminus for the complete unfolding of the  $J\alpha$ -helix is demonstrated by the mutational experiments of Harper *et al.*,<sup>12</sup> who

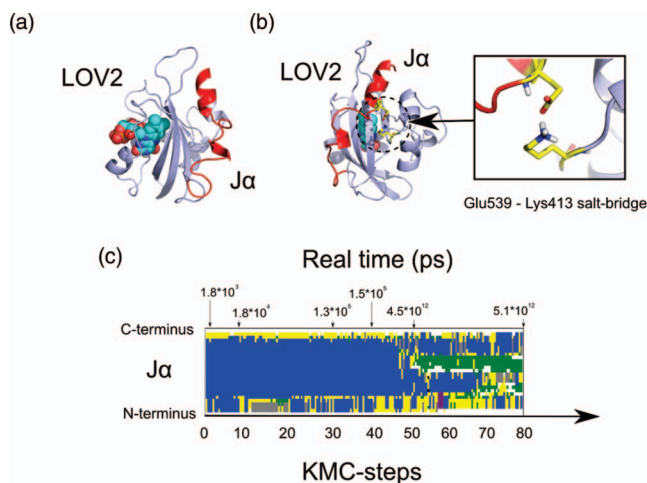


FIG. 5. (a) Side-view of the final tertiary structure, (b) alternative side-view of the final tertiary structure with snapshot of Glu539-Lys413 salt bridge, and (c) secondary-structure analysis of the J $\alpha$ -helix as a function of KMC steps and real time from Ile539Glu-AsLOV2 mutant in the dark state, obtained with KMC-MD method.

performed the point mutation Ile539Glu at the C-terminus of the J $\alpha$ -helix. To study the consequence of this mutation on the secondary structure of the J $\alpha$ -helix, we consider next in Fig. 5 the results of the KMC-MD simulation of the Ile539Glu-AsLOV2-J $\alpha$  mutant in the dark-state form. From the secondary structure analysis of the sequential range of the J $\alpha$ -helix, we deduce that at the 46th KMC step the J $\alpha$ -helix loses

its secondary structure, which is due to the formation of a salt bridge between Glu539-OE2 and the N-terminal Lys413-NZ (see snapshot in Fig. 5(b)). This is in agreement with the observations of Harper *et al.*,<sup>12</sup> who found that complete unfolding of the mutated J $\alpha$ -helix takes place even in the dark-state form. We conclude from these observations that, additionally to the detachment process of the J $\alpha$ -helix, our KMC-MD method reproduces the experimental findings about its subsequent unfolding in a reliable way. To study and understand in more detail the different disruption times, we compare in Fig. 6 the tertiary structures and APBS-electrostatic surfaces of the AsLOV2-J $\alpha$ -interfacial region and the J $\alpha$ -helix from the crystal structure of Halavaty and Moffat<sup>10</sup> as well as the NMR-solution structure of Harper *et al.*<sup>11</sup> We conclude from the tertiary structures of the AsLOV2-J $\alpha$  surfaces that the J $\alpha$ -helix in case of the crystal structure of Halavaty and Moffat<sup>10</sup> is longer and has a different orientation on the  $\beta$ -sheet of the AsLOV2 core than in case of the solution structure of Harper *et al.*<sup>11</sup> Moreover, we deduce from the APBS surfaces that the AsLOV2-J $\alpha$ -interfacial region and the J $\alpha$ -helix in the former case are less polar and, therefore, more hydrophobic, which leads to a reduced attractivity for water to enter the AsLOV2-J $\alpha$  interface and to a lower tendency to dissociate both entities. This causes that the AsLOV2-J $\alpha$  configuration from the crystal is thermodynamically more stable and, thus, has a higher disruption time than the one from solution. Finally, from Fig. 4(d) we can estimate the time for the dissociation of the J $\alpha$ -helix from the AsLOV2 core and its subsequent partial unfolding to take place in the range from 130 ns up to

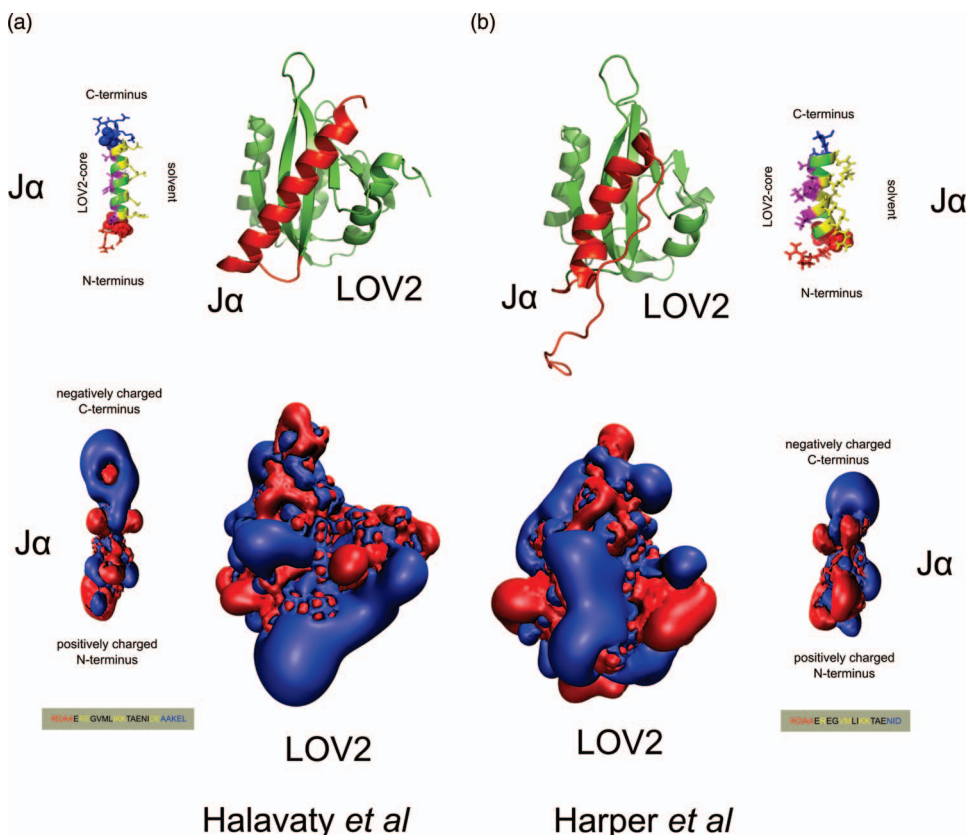


FIG. 6. (a) Front-view and APBS surface of the AsLOV2 core with the J $\alpha$ -helix from the x-ray crystal structure of Halavaty and Moffat.<sup>10</sup> (b) Front-view and APBS surface of the AsLOV2 core with the J $\alpha$ -helix from NMR-solution structure of Harper *et al.*<sup>12</sup>



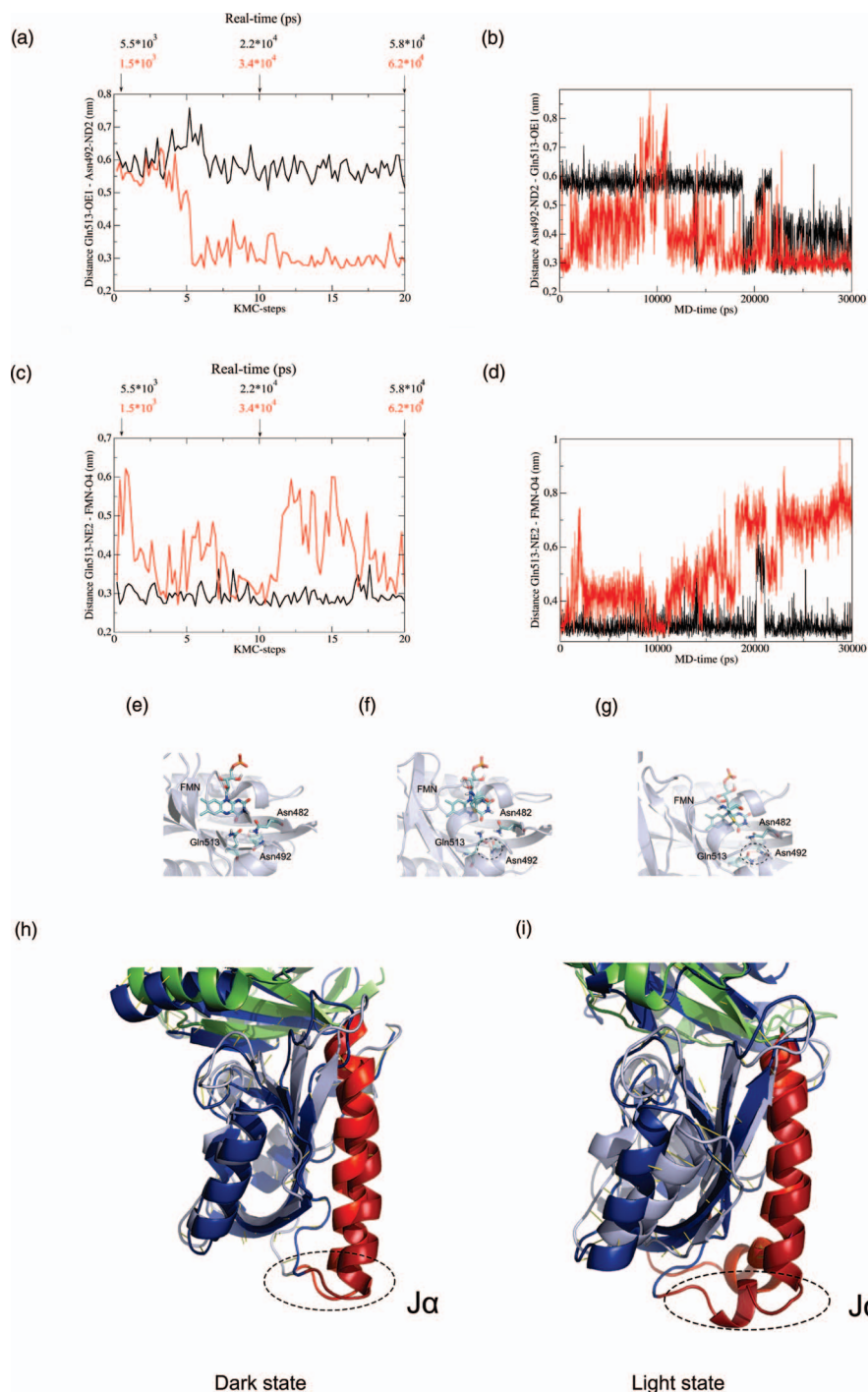


FIG. 7. Inter-atomic distances and representative structures of the PA-Rac1 system, obtained with KMC-MD- and non-invasive MD-methods. (a) Inter-atomic distance between Gln513-OE1 and Asn492-ND2 from KMC-MD simulation in the RB-KMC phase as a function of KMC steps. (b) Inter-atomic distance between Gln513-OE1 and Asn492-ND2 from MD simulation as a function of MD time. (c) Inter-atomic distance between Gln513-NE2 and FMN-C4=O from KMC-MD simulation in the RB-KMC phase. (d) Inter-atomic distance between Gln513-NE2 and FMN-C4=O from MD simulation as a function of MD time [red curve: light state; black curve: dark state]. Final representative amino-acid configurations near to the FMN chromophore, obtained from (e) KMC-MD simulation of the dark state, (f) KMC-MD simulation of the light state as well as (g) MD simulation of the light state [cyan: C-atom; red: O-atom; blue: N-atom; white: H-atom; orange: P-atom; yellow: S-atom]. Overlaid final structures of dark state (h) and light state (i) of PA-Rac1 system after 30 ns of MD- (dark colors) and 40 KMC-MD-steps (bright colors).

330  $\mu$ s. This can be correlated with the experimental time range of 300  $\mu$ s–1 ms for these events, determined by Nakasone *et al.*<sup>19</sup> for the LOV2 domain of phototropin1 from *Arabidopsis* using transient-grating methods. Thus, we conclude from the previous investigation that the KMC-MD method reproduces the multiscale relaxation behavior of the AsLOV2-

J $\alpha$  system after CFN-adduct formation reliably and permits to extend the computationally accessible simulation times to experimental timescales.

In the next part, we study the usefulness of the KMC-MD method to describe the multiscale relaxation dynamics of complex photoenzymes. As an example, we consider the



signaling behavior of the photoactivable Rac1 PA-Rac1 system, where the Rac1 enzyme has been fused with the AsLOV2- $J\alpha$  photoswitch. We start by analyzing the structural-dynamical changes in the RB-KMC phase of the amino acids in vicinity of the FMN chromophore, suspected to be involved in the primary steps of the signaling pathway of PA-Rac1 in previous MD simulations on the same system.<sup>16</sup> In Fig. 7(a), we plot the distances between Gln513-OE1 and Asn492-ND2 from the dark- and light-states, obtained with the KMC-MD method, as a function of KMC steps and real time. We see that after the 5th KMC step the light-state curve undergoes a severe drop from 0.6 nm to H-bond distance at around 0.3 nm, whereas the dark-state curve performs a stable fluctuation around the former value throughout the whole RB-KMC phase. Similar time behavior is observed for the quantity after an initial relaxation phase in MD simulations with the PA-Rac1 system, as shown in Fig. 7(b), as well as in KMC-MD simulations with the isolated AsLOV2- $J\alpha$  system in Fig. 2(a). Next, we compare in Figs. 7(c) and 7(d) the distances between Gln513-NE2 and the carbonyl-oxygen FMN-C4=O of the dark- and light-states, obtained with the KMC-MD- and MD-methods, respectively. We notice that, while in case of the dark state the curves remain stable at H-bond distance, the light-state curves fluctuate at increased distance with large magnitude. To further illustrate the changes in the amino-acid environment of the FMN chromophore, we show in Figs. 7(e) and 7(f) the respective representative snapshots from the final configurations of PA-Rac1 in its dark-state and light-state form after the RB-KMC phase in comparison to the final configuration of PA-Rac1 in the light-state from MD simulation in Fig. 7(g). From these graphs, we deduce that after CFN-adduct formation the Gln513, located on the  $I\beta$ -strand, detaches from the carbonyl-oxygen FMN-C4=O and forms a H-bond with the adjacent Asn492 on the  $H\beta$ -strand, which results in the coupling between the  $I\beta$ - and  $H\beta$ -strands. We point out that this process has been found by us to be crucial in triggering the signaling pathways of the AsLOV2- $J\alpha$ - and Vivid-LOV-photoswitches, as demonstrated previously and in the MD-simulation studies of Refs. 14 and 54. We conclude from our investigation that it is also well reproduced by our RB-KMC-simulation algorithm, introduced in Sec. II A 1. To assess the consequences of the triggering process near the FMN chromophore on the tertiary structure of the PA-Rac1 system, we further visualize in Figs. 7(h) and 7(i) the final configurations of the LOV2- $J\alpha$  domain from PA-Rac1 in its dark-state and light-state form, respectively, obtained after the ID-KMC phase and MD simulation. We deduce from these graphs that the coupling between the  $I\beta$ - and  $H\beta$ -strands causes a N-terminal disruption of the  $J\alpha$ -helix in the light state of the PA-Rac1 system, which is not observed in the dark state. Next, we analyze the impact of this disruption on the functional activity of the Rac1 enzyme. To this end, we display in Figs. 8(a) and 8(b) the final structures from the ID-KMC phase of PA-Rac1 in the dark- and light-states, respectively. From Fig. 8(b), we deduce that in the light state the  $A\beta$ - $B\beta$ -loop detaches from the activation site switchII on the GTPase, which is triggered by the N-terminal disruption of the  $J\alpha$ -helix. This causes that the steric inhibition from the LOV2 core is relieved and the activation site of

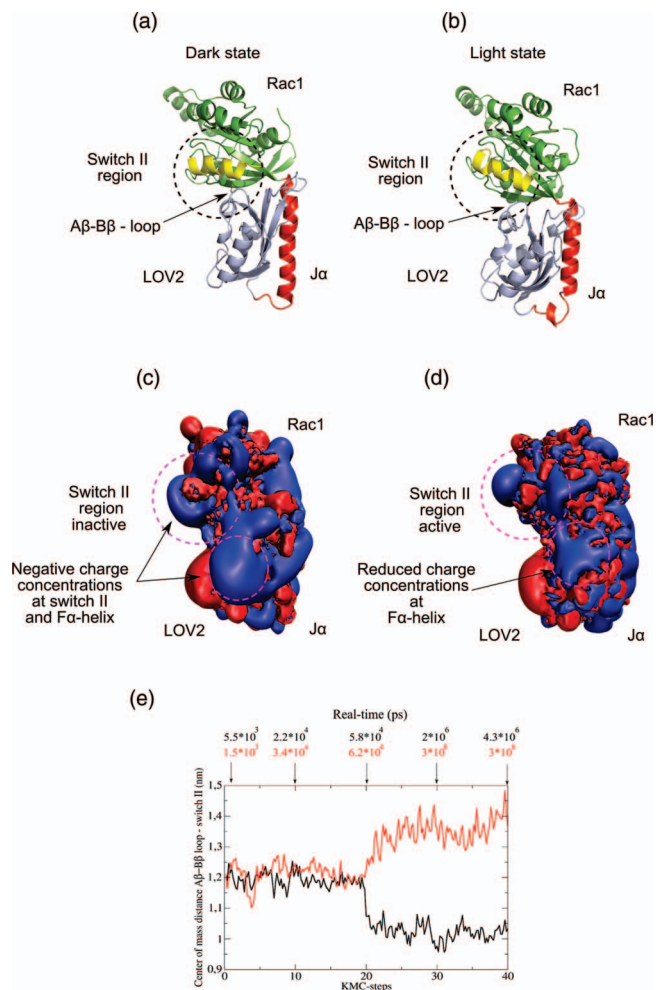


FIG. 8. Final structures of PA-Rac1 system in the dark state (a) and light state (b) from KMC-MD simulation after 40 KMC-MD steps. APBS surfaces of PA-Rac1 system in the dark state (c) and light state (d). (e) Center-of-mass distance between  $A\beta$ - $B\beta$  loop and switchII from KMC-MD simulation of the light state (red curve) and dark state (black curve) as a function of KMC steps.

Rac1 becomes accessible to interactions with its effector domains, such as PAK1.<sup>74-76</sup> By contrast, we infer from Fig. 8(a) that in case of the dark state the switchII region is blocked by the  $A\beta$ - $B\beta$ -loop, which causes that the Rac1 enzyme is inactive under dark-state conditions. To elucidate the reasons for this behavior, we consider in Figs. 8(c) and 8(d) the APBS-electrostatic surfaces from the final dark-state- and light-state structures of PA-Rac1, respectively. In case of the dark state we see that the switchII region as well as the  $A\beta$ - $B\beta$  loop possess a high negative electrostatic surface charge, which causes that both structural elements repel each other and are subjected to a high tension. This tension is released upon illumination through the N-terminal disruption of the  $J\alpha$ -helix from the LOV2 core, as can be concluded from the reduced charge concentrations at the same sites observed in Fig. 8(d). Ultimately, this leads to the detachment of the  $A\beta$ - $B\beta$  loop from the switchII region, which is confirmed by the increase in the distance of the respective centers of mass of the light state visualized in Fig. 8(e). From the same figure, we deduce that in the final stage of the KMC-MD simulation the

difference in the distance between these two structural elements in the dark- and light-state amounts to 0.4 nm, which remains stable over several microseconds. We emphasize that these observations are in agreement with the x-ray measurements of Wu *et al.*,<sup>15</sup> who demonstrated that the A $\beta$ -B $\beta$  loop from the AsLOV2 domain in PA-Rac1 acts as a steric inhibitor enabling the light-sensitive regulation of the functional activity of the Rac1 enzyme.

To conclude, we evaluate the computational expense required by the KMC-MD method in comparison to the MD-simulation technique, to access similar timescales. To this end, let us consider the light state of the PA-Rac1 system, for which we reached a total simulation time of 300  $\mu$ s using the KMC-MD method. In this calculation a total number of 40 KMC steps were accomplished, requiring a total number of 200 000 MD-relaxation steps. Now, if we would use the conventional MD-simulation technique with a time step of 1 fs, we would need a total number of  $3 \times 10^{11}$  MD steps, to reach the same time of 300  $\mu$ s as obtained with the KMC-MD approach. By considering that the costs to execute the KMC events are on the order of a MD-relaxation phase, we conclude that the KMC-MD-simulation method reduces the computational expense by a factor of  $7.5 \times 10^5$  for the PA-Rac1 system in the light state.

#### IV. CONCLUSIONS

In summary of this work we introduced a novel multiscale-modeling method, based on a combination of the kinetic Monte Carlo- and molecular-dynamics-technique, and demonstrated its usefulness for the simulation of the multiscale signaling behavior of the AsLOV2-J $\alpha$  photoswitch as well as the AsLOV2-J $\alpha$ -regulated photoactivable Rac1-GTPase. After photoexcitation with blue light, these systems typically exhibit a complex signaling pathway by carrying out multiscale stress relaxation dynamics. This implicates local structural changes at the residue level near to a light-sensitive reaction center on a nanosecond timescale, inducing extensive structural rearrangements between peripheral protein domains on timescales ranging from several microseconds up to seconds. More specifically, with our new approach we find that the early stages of the signaling pathway are characterized by the coupling between the I $\beta$ - and H $\beta$ -strands through H-bond formation between Gln513 and Asn492. This event leads to a N-terminal unfolding of the J $\alpha$ -helix and its subsequent disruption from the LOV core. In case of the photoactivable Rac1 system, we show that this latter disruption process results in the release of a functional  $\alpha$ -helix located at the switchII region of the Rac1 enzyme, which is known to be the primary docking site for Rac1 with its effector domains, such as PAK1. Through comparing these findings with the available experimental results, we conclude that our method is a powerful new multiscale-modeling tool for simulating the signaling behavior of large protein complexes from ultrafast up to typical timescales encountered in biological experiments. Further gains in efficiency can be expected by using multiple-timestepping approaches, such as RESPAs, to accelerate the calculation of the MD-relaxation phase. Finally, we emphasize that its underlying algorithm is trans-

ferable to a large number of alternative protein-solvent systems and will open new avenues for investigating the light-induced regulation of enzymatic reactivity<sup>77</sup> as well as gene function.<sup>78</sup>

#### ACKNOWLEDGMENTS

The work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the Research Training Group GRK 640 "Sensory photoreceptors in natural and artificial systems" as well as the Research Training Group GRK 1626 "Chemical photocatalysis."

- <sup>1</sup>J. B. Pereira-Leal, E. D. Levy, and S. A. Teichmann, *Philos. Trans. R. Soc. London Ser. B* **361**, 507 (2006).
- <sup>2</sup>K. Wennerberg and C. J. Der, *J. Cell Sci.* **117**, 1301 (2004).
- <sup>3</sup>A. Singh, A. E. Karnoub, T. R. Palmby, E. Lengyel, J. Sondek, and C. J. Der, *Oncogene* **23**, 9369 (2004).
- <sup>4</sup>P. Hegemann, *Annu. Rev. Plant Biol.* **59**, 167 (2008).
- <sup>5</sup>J. M. Christie, *Annu. Rev. Plant Biol.* **58**, 21 (2007).
- <sup>6</sup>W. R. Briggs, *J. Biomed. Sci.* **14**, 499 (2007).
- <sup>7</sup>T. Kottke, P. Hegemann, B. Dick, and J. Heberle, *Biopolymers* **82**, 373 (2006).
- <sup>8</sup>M. A. Jones, K. A. Feeney, S. M. Kelly, and J. M. Christie, *J. Biol. Chem.* **282**, 6405 (2007).
- <sup>9</sup>D. Matsuoka and S. Tokutomi, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13337 (2005).
- <sup>10</sup>A. S. Halavaty and K. Moffat, *Biochemistry* **46**, 14001 (2007).
- <sup>11</sup>S. M. Harper, L. C. Neil, and K. H. Gardner, *Science* **301**, 1541 (2003).
- <sup>12</sup>S. M. Harper, J. M. Christie, and K. H. Gardner, *Biochemistry* **43**, 16184 (2004).
- <sup>13</sup>S. Crosson and K. Moffat, *Plant Cell* **14**, 1067 (2002).
- <sup>14</sup>E. Peter, B. Dick, and S. A. Baeurle, *Nat. Commun.* **1**, 122 (2010).
- <sup>15</sup>Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, and K. M. Hahn, *Nature (London)* **461**, 104 (2009).
- <sup>16</sup>E. Peter, B. Dick, and S. A. Baeurle, *Proteins: Struct., Funct., Bioinf.* (2012).
- <sup>17</sup>M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679 (2005), and references therein.
- <sup>18</sup>S. A. Baeurle, *J. Math. Chem.* **46**, 363 (2009).
- <sup>19</sup>Y. Nakasone, T. Eitoku, D. Matsuoka, S. Tokutomi, and M. Terazima, *J. Mol. Biol.* **367**, 432 (2007).
- <sup>20</sup>W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Biller, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren, *J. Phys. Chem. A* **103**, 3596 (1999).
- <sup>21</sup>L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, *J. Chem. Theory Comput.* **4**, 819 (2008).
- <sup>22</sup>A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. de Vries, and S. J. Marrink, *J. Comput. Chem.* **31**, 1333 (2010).
- <sup>23</sup>H. M. Senn and W. Thiel, *Top Curr. Chem.* **268**, 173 (2007).
- <sup>24</sup>J. Noolandi, T. S. Davison, A. R. Völkel, X.-F. Nie, C. Kay, and C. H. Arrowsmith, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9955 (2000).
- <sup>25</sup>A. R. Völkel and J. Noolandi, *Biophys. J.* **80**, 1524 (2001).
- <sup>26</sup>T. Coussaert, A. R. Völkel, J. Noolandi and A. P. Gast, *Biophys. J.* **80**, 2004 (2001).
- <sup>27</sup>G. S. Ayton and G. A. Voth, *Curr. Opin. Struct. Biol.* **19**, 138 (2009), and references therein.
- <sup>28</sup>M. Tuckerman, B. J. Berne, and G. J. Martyna, *J. Chem. Phys.* **97**, 1990 (1992).
- <sup>29</sup>C. R. Sweet, P. Petrone, V. S. Pande, and J. A. Izaguirre, *J. Chem. Phys.* **128**, 145101 (2008).
- <sup>30</sup>J. A. Izaguirre, C. R. Sweet, and V. S. Pande, *Pac. Symp. Biocomput.* **15**, 240 (2010).
- <sup>31</sup>D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, *Science* **330**, 341 (2010).
- <sup>32</sup>V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, *J. Am. Chem. Soc.* **132**, 1526 (2010).
- <sup>33</sup>M. Shirts and V. S. Pande, *Science* **290**, 1903 (2000).

- <sup>34</sup>D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, in *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)* (ACM, New York, 2009), pp. 1–11.
- <sup>35</sup>J. A. Izaguirre and S. S. Hampton, *J. Comput. Phys.* **200**, 581 (2004).
- <sup>36</sup>H. Grubmueller, *Phys. Rev. E* **52**, 2893 (1995).
- <sup>37</sup>A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, and M. Parrinello, *J. Am. Chem. Soc.* **128**, 2705 (2006).
- <sup>38</sup>A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- <sup>39</sup>D. E. Makarov and H. Metiu, *J. Chem. Phys.* **116**, 5205 (2002).
- <sup>40</sup>D. E. Makarov, C. A. Keller, K. W. Plaxco, and H. Metiu, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3535 (2002).
- <sup>41</sup>O. Collet, *Phys. Rev. E* **67**, 061912 (2003).
- <sup>42</sup>D. E. Makarov, P. K. Hansma, and H. Metiu, *J. Chem. Phys.* **114**, 9663 (2001).
- <sup>43</sup>P.-C. Li and D. E. Makarov, *J. Chem. Phys.* **119**, 9260 (2003).
- <sup>44</sup>M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109 (1997).
- <sup>45</sup>A. F. Voter, "Introduction to the kinetic Monte Carlo method," in *Radiation Effects in Solids*, edited by K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga (Springer, Berlin, 2007), Vol. 235, pp. 1–23.
- <sup>46</sup>J. Howard, "Mechanics of motor proteins," in *Physics of Bio-Molecules and Cells, 75th Les Houches Summer School*, edited by H. Flyvbjerg, F. Jülicher, P. Ormos, and F. David (Springer-Verlag, Berlin, 2002), Vol. 75, pp. 69–94.
- <sup>47</sup>A. A. Gurtovenko and Y. Y. Gotlib, *J. Chem. Phys.* **115**, 6785 (2001).
- <sup>48</sup>V. B. P. Leite, J. N. Onuchic, G. Stell, and J. Wang, *Biophys. J.* **87**, 3633 (2004).
- <sup>49</sup>P. L. Freddolino, M. Dittrich, and K. Schulten, *Biophys. J.* **91**, 3630 (2006).
- <sup>50</sup>T. Kottke, J. Heberle, D. Hehn, B. Dick, and P. Hegemann, *Biophys. J.* **84**, 1192 (2003).
- <sup>51</sup>R. Fedorov, I. Schlichting, E. Hartmann, T. Domratcheva, M. Fuhrmann, and P. Hegemann, *Biophys. J.* **84**, 2474 (2003).
- <sup>52</sup>E. Peter, B. Dick, and S. A. Baeurle, "Signals of LOV1: a computer simulation study on the wildtype LOV1-domain of *Chlamydomonas reinhardtii* and its mutants," *J. Mol. Model.* (to be published).
- <sup>53</sup>E. Peter, B. Dick, and S. A. Baeurle, *J. Chem. Biol.* **4**, 167 (2011).
- <sup>54</sup>E. Peter, B. Dick, and S. A. Baeurle, *Proteins: Struct., Funct., Bioinf.* **80**, 471 (2012).
- <sup>55</sup>K. J. Laidler and M. C. King, *J. Phys. Chem.* **87**, 2657 (1983).
- <sup>56</sup>D. K. Roylance, "Characterization of polymer deformation and fracture," in *Applications of Polymer Spectroscopy*, edited by E. G. Brame (Academic, New York, 1978), pp. 207–219.
- <sup>57</sup>T. Keii, *Heterogeneous Kinetics: Theory of Ziegler-Natta-Kaminsky Polymerization* (Springer Series in Chemical Physics, Berlin, 2004), pp. 59–60.
- <sup>58</sup>D. C. Liu and J. Nocedal, *Math. Program* **45**, 503 (1989).
- <sup>59</sup>D. Fincham, N. Quirke, and D. J. Tildesley, *J. Chem. Phys.* **84**, 4535 (1986).
- <sup>60</sup>See supplementary material at <http://dx.doi.org/10.1063/1.3697370> for the process pathes of the H-bond breakage, H-bond formation as well as side chain rotation events. In addition detailed information is provided about the determination of the pre-exponential factors for the KMC-MD simulations of the AsLOV2-J $\alpha$ - and PA-Rac1-systems.
- <sup>61</sup>A. Luzar and D. Chandler, *J. Chem. Phys.* **98**, 8160 (1993).
- <sup>62</sup>B. Hetenyi, F. De Angelis, P. Giannozzi, and R. Car, *J. Chem. Phys.* **120**, 8632 (2004).
- <sup>63</sup>I. W. Kuo and C. J. Mundy, *Science* **303**, 658 (2004).
- <sup>64</sup>Y. Zhang, M. Lagi, D. Liu, F. Mallamace, E. Fratini, P. Baglioni, E. Mamontov, M. Hagen, and S.-H. Chen, *J. Chem. Phys.* **130**, 135101 (2009).
- <sup>65</sup>E. Lindahl, B. Hess, and D. van der Spoel, *J. Mol. Model.* **7**, 306 (2001).
- <sup>66</sup>T. Soares, X. Daura, C. Oostenbrink, L. Smith, and W. Gunsteren, *J. Biomol. NMR* **30**, 407 (2004).
- <sup>67</sup>N. Todorova, F. S. Legge, H. Treutlein, and I. Yarovsky, *J. Phys. Chem.* **112**, 11137 (2008).
- <sup>68</sup>D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 2003).
- <sup>69</sup>C. Neiss and P. Saalfrank, *Photochem. Photobiol.* **77**, 101 (2003).
- <sup>70</sup>A. I. Nash, W.-H. Ko, S. M. Harper, and K. H. Gardner, *Biochemistry* **47**, 13842 (2008).
- <sup>71</sup>M. T. A. Alexandre, R. van Grondelle, K. J. Hellingwerf, and J. T. M. Kennis, *Biophys. J.* **97**, 238 (2009).
- <sup>72</sup>S. Arai, M. Togashi, M. Shiozawa, Y. Inoue, and M. Sakurai, *Chem. Phys. Lett.* **414**, 230 (2005).
- <sup>73</sup>A. Pfeifer, T. Majerus, K. Zikihara, D. Matsuoka, S. Tokutomi, J. Heberle, and T. Kottke, *Biophys. J.* **96**, 1462 (2009).
- <sup>74</sup>C. DerMadirossian, A. Schnelzer, and G. M. Bokoch, *Mol. Cell* **15**, 117 (2004).
- <sup>75</sup>D. Owen, L. J. Campbell, K. Littlefield, K. A. Evetts, Z. Li, D. B. Sacks, P. N. Lowe, and H. R. Mott, *J. Biol. Chem.* **283**, 1692 (2008).
- <sup>76</sup>J. Yamauchi, Y. Miyamoto, A. Sanbe, and A. Tanoue, *Exp. Cell Res.* **312**, 2954 (2006).
- <sup>77</sup>A. Deiters, *ChemBioChem* **11**, 47 (2010).
- <sup>78</sup>A. Deiters, *Curr. Opin. Chem. Biol.* **13**, 678 (2009).