MCMap—A Computational Tool for Mapping Energy Landscapes of Transient Protein–Protein Interactions

Johannes M. Foerster,†‡ Ina Poehner,†‡ and G. Matthias Ullmann*‡

Computational Biochemistry, University of Bayreuth, Universitätsstr. 30, NW1, 95447 Bayreuth, Germany

ABSTRACT: MCMap is a tool particularly well-suited for analyzing energy landscapes of transient macromolecular complexes. The program applies a Monte Carlo strategy, where the ligand moves randomly in the electrostatic field of the receptor. By applying importance sampling, the major interaction sites are mapped, resulting in a global distribution of ligand–receptor complexes. This approach displays the dynamic character of transiently interacting protein complexes where not a single complex but an ensemble of complexes better describes the protein interactions. The software provides a broad range of analysis options which allow for relating the simulations to experimental data and for interpreting them on a structural level. The application of MCMap is exemplified by the electron-transfer complex of cytochrome c peroxidase and cytochrome c from baker’s yeast. The functionality of MCMap and the visualization of simulation data are in particular demonstrated by studying the dependence of the association on ionic strength and on the oxidation state of the binding partner. Furthermore, microscopically, a repulsion of a second ligand can be seen in the ternary complex upon the change of the oxidation state of the bound cytochrome c. The software is made available as open source software together with the example and can be downloaded free of charge from http://www.bisb.uni-bayreuth.de/index.php?page=downloads.

INTRODUCTION

Protein–protein interactions play a central role in many cellular processes such as signal transduction, gene regulation, and molecular bioenergetics. The formed protein complexes can be very stable with low dissociation rates as, for instance, in the case of antigen–antibody complexes, or they can be very short-living, having only a transient nature, as, for instance, in the case of electron-transfer protein complexes. Especially for electron-transfer protein complexes, it is important to find a reasonable balance between directional association to ensure correct pairing of partners and the possibility of quick exchange. Thus, the interactions between the electron-transfer proteins are often very dynamic compared to other complexes. In many cases, a single complex structure cannot describe the diversity of complexes for transiently interacting proteins very well but can be better described by an ensemble of complexes assuming different orientations and binding sites. This behavior can also be observed in crystal structures of several electron-transfer complexes, where the electron density of the smaller protein is often not as well defined as that of the receptor. Multiple binding conformations and orientations of the ligand result in a structure ensemble, which causes a more or less smeared electron density of the smaller binding partner. A similar observation was made before, where a combination of specific and loosely bound complexes was used to describe the diversity of an encounter complex. The combination of replica exchange Monte Carlo (MC) simulations and experimental paramagnetic relaxation enhancement (PRE) measurements revealed that certain key residues often account for protein contacts in both, the specific and unspecific ones. Transient complexes have usually high dissociation rates, and the proteins often have only minor surface complementarity. Protein association processes for these types of proteins can be classified in two to three association steps. If the two molecules are far away from each other, mainly long-range interactions such as electrostatic interactions are important for attractive forces bringing them in the vicinity of each other, forming the so-called encounter complex. When the proteins are in contact distance, they can scan the surfaces of each other. This phase of the association is often described by an ensemble of energetically favorable orientations. During this phase, short-range interactions such as hydrogen bonds and hydrophobic

Received: March 26, 2018
Accepted: May 22, 2018
Published: June 18, 2018

DOI: 10.1021/acsomega.8b00572
ACS Omega 2018, 3, 6465−6475

© 2018 American Chemical Society
interactions become more and more important and contribute to the complex formation and stabilization of the final complex. In order to get a first idea of binding, mainly long-range interactions need to be considered. Interactions such as electrostatic attraction and repulsion mainly contribute to the primal complex formation and direct the proteins toward the binding areas. The theoretical analysis of protein–protein association is a major challenge, especially for transient protein complexes. Many different approaches are commonly used. An example of a widely used tool for evaluating binding affinities to a certain extent is docking simulations. For many docking tools, it is, however, a major challenge to predict yet unknown binding sites. Different tools have been applied for finding and validating first guesses of docking sites, for example, molecular dynamics simulations, Brownian dynamics, and MC dockings involving random movements. In previous studies, we could demonstrate that such MC predictions can be well-correlated with experimental data, namely, with NMR studies, in which spin labels causing PREs have been attached to the proteins in order to visualize the lowly populated states of the encounter ensemble of transient complexes. The simulated encounter ensemble reflected the dynamic behavior of the transient complexes well, which cannot be described by a single protein complex.

In this paper, we present the background and usage of the program MCMap, which approximates the distribution of possible protein interactions by generating an ensemble of protein complexes (distributed over the surface of a receptor). The program uses a Metropolis MC algorithm to evaluate the interaction energy and, with this, the probability of an encounter to occur. It can be used as a tool to identify binding sites in a rigid-body docking approach, especially if the association is driven by electrostatic interactions, which is often the case for protein–protein complexes. Hydrophobic interactions are currently not included in the energy calculation. The inclusion of these interactions is, however, planned for future development of MCMap. As a ligand, a small molecule, another protein, or a nucleic acid can be used. For demonstrating the usage and the capabilities of the software, we analyze the well-studied complex of cytochrome c peroxidase (CcP) and cytochrome c (Cc) from Saccharomyces cerevisiae. CcP and Cc are located in the intermembrane section of mitochondria, where the ferrous form of Cc forms a complex with CcP to reduce hydrogen peroxide to water. The complex formation of the two electron-transfer proteins is mainly driven by electrostatics. Thus, long-range charge–charge attraction allows for preorientation of proteins and thus limits the conformational search for the final complex formation. Recently, it was shown that not only long-range interactions play a role in protein association but also the local interactions can have some impact on the association process of two proteins. In the first section of the paper, we introduce the algorithm used in MCMap and the methods for analyzing the simulation results. In the second part, we illustrate the analysis of the complex distributions obtained by MCMap. In particular, we analyze the association of CcP and Cc in different oxidation states and at different ionic strength values. The software and the source code are made available free of charge and can be downloaded from http://www.bisb.uni-bayreuth.de/index.php?page=downloads.

**SOFTWARE FEATURES**

**MC Simulation of Association.** The program MCMap generates a distribution of encounter complexes by randomly translating and rotating one molecule in the field of the other using Metropolis MC. This ensemble characterizes the transient complex and possible interactions in this complex. The density of complexes resembles the probability of an encounter at the specific position. Because the program samples all translational and rotational degrees of freedom, MCMap can even be applied if little is known about the binding site. The protein structures of the two partners, which we, from now on, refer to as the ligand and the receptor, are kept fixed during the MC simulation while translational and rotational movements are allowed. A general overview of the workflow of MCMap can be seen in Figure 1. For computational reasons, it is advantageous to use the smaller docking partner as the ligand. The structures have to be provided in the PQRM-format, which is a modified PDB-format containing partial charges, radii, and masses of the atoms in the last three columns after the xyz coordinates. Such a PQRM-file can, for instance, be obtained from CHARMM results with the help of the converter program psfcrd2pqr, which is provided together with MCMap. The electrostatic potentials are provided as OpenDX-files (a format for representing volumetric data) that can, for instance, be calculated by the program APBS, which solves the Poisson–Boltzmann equation for a heterogeneous dielectric environment. The electrostatic potential is provided on two grids, a large coarse grid and a small fine grid. The fine grid allows a more detailed representation of the electrostatic potential when the two proteins are in closer proximity. The overall flowchart of the MC algorithm is shown in Figure 2. A schematic representation of an MCMap run is depicted in Figure 3. In the beginning of a simulation, the

![Figure 1. Overall workflow diagram for the usage of MCMap: MCMap needs structural information, the electrostatic potentials, and an exclusion grid as a starting point. After simulating the association, the results are processed with print-coor. Depending on the chosen analysis, the simulation results can be visualized with the focus on either the receptor or the ligand. Alternatively, the whole simulation can be analyzed by contact map histograms.](image-url)
Φ

lig and the receptor is obtained by multiplying the charges $q_{n_{lig}}$ of the ligand with the electrostatic potential $\Phi_{rec}(r)$ of the receptor (eq 1).

$$E_{int} = \sum_{n=0}^{N_{lig}} q_{n_{lig}} \Phi_{rec}(r_n)$$

where $N_{lig}$ is the number of atoms of the ligand. The potential $\Phi_{rec}$ at the position $r_n$ of the charge $d_{lig}$ is obtained by linear interpolation of the surrounding grid points. Subsequently, the ligand is randomly translated and rotated, and the energy is re-evaluated using eq 1. This random move is accepted or rejected according to the Metropolis criterion. The next steps in the simulation are performed equivalently. In order to allow for a more detailed sampling of the interaction surface in the vicinity of the proteins, the MC moves are scaled down. In each MC move, it is checked whether the ligand reaches a specified distance $R_{out}$ originating from the center of mass of the receptor. If the ligand reaches this distance, it is being considered to be too far away from the receptor for a proper interaction and the ligand is reinitialized at the starting surface with the radius $R_{in}$. This approach shows clear similarities to the Brownian dynamics procedure but has the advantage that no forces need to be calculated and the generated ensemble directly represents a Boltzmann ensemble without further data manipulation.

In addition to the evaluation of the electrostatic energy, it has to be made sure that the two molecules do not penetrate each other. The usage of a Lennard-Jones type of energy term would be computationally too expensive because it would require $N_{lig} \times N_{rec}$ distance evaluations. Instead, we use an exclusion grid, which is a regular grid describing the shape of the receptor. The exclusion grid is generated by the program make-excl and saved as a binary file. Each point of the exclusion grid is flagged being either in the receptor or in the solvent. If, during the simulation, the distance between the ligand and the receptor is such that the two molecules can collide, the exclusion grid is invoked. Thereby, it is tested if a surface atom of the ligand is entering a grid cell of the exclusion grid that is occupied by the receptor. If that is the case, the move is rejected. The test is computationally very inexpensive and requires only a few rounding processes to identify the appropriate grid points. Moreover, the number of operations just scales linearly with the number of surface atoms of the ligand. In case of several consecutive rejections, it is assumed that the simulation got stuck in a local minimum and the ligand is reinitialized at the starting radius $R_{in}$.

The MC simulation proceeds until the maximum number of MC runs is reached. The ensemble is stored in an orientation file, which contains information on the energy and the coordinate transformation required for generating the orientation from the initial structure. Each line of this orientation file, which is a compressed ASCII file, contains one orientation. Storing the ensemble in this form has several advantages. Besides saving storage space compared to a storage of full structures, it is also possible to use only a subset of the original structure, for instance, only the prosthetic groups, for the generation of the orientation. This method substantially speeds up the analysis. Moreover, this orientation file can, for instance, be resorted by energy, facilitating the use of only orientations with energies below a certain threshold for further investigation. Thus, the way how MCMAP saves its generated ensemble gives a large flexibility for the subsequent analysis.

The runtime of the simulation does not depend on the size of the receptor but mainly depends on the number of atoms in the ligand. Hence, it is recommended to use the smaller center of mass of the ligand is randomly placed on the surface of a sphere with the radius $R_{in}$. The center of mass of the receptor is placed at the center of this sphere. The electrostatic potential of the receptor at this sphere should be very close to zero or at least have a spherical symmetry to a good approximation in order to avoid a bias in the calculation stemming from the initialization. The interaction energy $E_{int}$ between the ligand and the receptor is obtained by multiplying the charges $q_{n_{lig}}$ of the ligand with the electrostatic potential $\Phi_{rec}(r)$ of the receptor (eq 1).
molecule as a ligand to speed up the simulation. In our example, a MCMAP simulation with a ligand of 1772 atoms and $10^5$ MC runs, each run consisting of $2.5 \times 10^4$ MC steps, results in about $2.6 \times 10^4$ different orientations and takes approximately 20 min on a single core Intel Core i7-4790 with a speed of 3.6 GHz.

**Ensemble Analysis and Processing.** The MCMAP suite provides a couple of tools to analyze the generated ensemble of configurations. The main utility for processing the distribution of configurations is the program print-coor. This program can, on the one hand, be used for generating representations of the ensemble, which can be visualized with molecular viewers such as VMD or PyMOL. On the other hand, print-coor can reduce the encounter ensemble to highlight regions of interest. In the way MCMAP stores the results, that is, representing each structure of the ensemble by a linear transformation of the reference orientation, the distribution of ensemble structures can be inspected not only with respect to the receptor but also vice versa with respect to the ligand. This possibility allows for the easy identification of high-affinity binding patches on the surface of either the ligand or the receptor.

The program print-coor can analyze the ensemble contained in the orientation file of MCMAP in different modes. One possibility is to generate each individual orientation of the ligand or of the receptor, which allows for analyzing individual orientations directly. Another way to analyze the ensemble of configurations is to generate the center of mass of either the ligand around the receptor or the receptor around the ligand. This information can also be visualized as a density (in OpenDX format). This density gives a fast overview of the simulation results and can be inspected with many molecular viewers in different representations such as isocontour plots or volume density slices. An example for such representations can be seen in Figure 4.

The program print-coor can not only help to visualize the structure ensemble but also reduce it as follows. Print-coor creates subensembles, which make a deeper inspection of specific regions easier. For this purpose, the program determines the distance between the ligand and the receptor and only orientations in which the distance between any given atom of the ligand and receptor is smaller than a given threshold $d$ are taken into account for this subensemble. Another way to alter the ensemble is to invoke an inclusion grid. The idea thereby is similar to the procedure for excluding structures that case of protein collisions, as described in the section of the MCMAP algorithm. Instead of excluding a ligand orientation entering the region that is occupied by the receptor, all orientations entering the regions that are defined by an inclusion grid are saved in a new subensemble. The inclusion grid presents a more efficient way for estimating whether certain regions of ligand and receptor are in proximity of each other upon complex formation. Similar to the exclusion grid, the inclusion grid can be generated by the program make-excl; the probe sphere radius needs to be adjusted to include larger regions around the receptor. This approach is computationally efficient and particularly of interest if the receptor molecule is nonspherical.

An overview of the simulations can also be obtained by creating a contact map histogram of all orientations, allowing for mapping major interaction regions to single atoms or amino acids. If the distance of an amino acid or atom of the ligand to an amino acid or atom of the receptor is smaller than a specified threshold, the counter in the 2D histogram is increased. The resulting histogram is written as an ASCII file, which can be plotted with PyCoALA (the Python Contact Area Localization and Analysis tool, distributed with MCMAP) or any plotting tool such as GMT or alternatively be further processed. For relating histograms of different simulations to each other, each histogram is normalized to the total amount of orientations saved in the individual MC simulation. This normalization, which prevents misinterpretation by comparing simulations resulting in ensembles which are differing in size, can be easily performed with PyCoALA. For a better representation, the histogram is then scaled such that the maximal occurrence is set to one.

Together with MCMAP, we provide PyCoALA, i.e., a tool that supports a representation of a contact map histogram on the protein structure using two different modes, namely, (i) hot spot visualization or (ii) difference mapping by histogram subtraction. For both types of modes, contact map histograms are used as inputs, which can have either atomic or amino acid resolution. The contact surface regions of interest are represented on the respective protein surface. For this purpose, the residues that are involved in more than a certain number of contacts are colored on the surface of the receptor and ligand by a PyMOL script. If differences between various simulations need to be examined, a subtraction of the results can help to identify the areas where the major differences occur. Positive values of the subtracted data are visualized in blue, and negative values are visualized in red. Analogous to the heatmap representation, the contact differences can be mapped back on the protein surface. An example for a visualization with the PyCoALA back-mapping strategy is described in the example below.

Figure 4. Overview of different representations of the center of mass density obtained from the MCMAP runs. CcP is represented with the electrostatic potential mapped on the surface ranging from $-5$ (red) to $5\text{ kcal/}e\text{V}$ (blue). Cc is shown in silver, with the heme being highlighted in red. The green isosurface represents the positions where the center of mass of Cc was found at least 100 times in the ensemble. The blue color in the background illustrates a slice through the volume of the obtained ensemble ranging from 0 (white) to at least 40 (dark blue) orientations.
In this section, we demonstrate the usage of the MCMap suite by applying the software to the complex of CcP and Cc. This complex has been analyzed in many previous experimental and also some theoretical studies. Moreover, structurally, this complex was characterized by X-ray crystallography and NMR. The complex shows many interesting features. Most notably, its association is largely governed by electrostatic interactions. Consequently, the association depends strongly on the ionic strength of the solution. Moreover, because the two reaction partners perform an electron-transfer reaction, their association is influenced by the redox state of the reaction partners. In the following parts, we will demonstrate how the influence of these different parameters can be simulated, analyzed, and visualized using MCMap.

**Influence of Ionic Strength on the Association.** The complex formation of CcP and Cc depends strongly on electrostatic interactions and is therefore well-suited to be studied with MCMap. In order to determine how protein association is influenced by the salt concentrations, we studied complex formation at 0, 10, 20, 50, and 100 mM ionic strength. A change in the ionic environment can result in a relocation of binding sites. An overview of such a docking analysis for different ionic strength values is shown in Figure 5. At a lower ionic strength, two binding sites can be seen, a major one at a negatively charged surface area of CcP and a minor one at the site distal from the primary site. Inspection of the density of Cc around CcP and vice
versa shows that the specific interaction surface of Cc shrinks with increasing ionic strength, causing random encounters to be more populated at higher ionic strength.

Because CcP and Cc are electron-transfer proteins, the distance between the redox cofactors, namely, the hemes, is an important parameter that can be analyzed in the generated ensemble. If one plots the probability of occurrence of encounter complexes in dependence of the shortest heme-to-heme distance of CcP and Cc and the corresponding energy, two minima can be identified at low ionic strength. These two minima represent two preferred binding orientations. With increasing ionic strength, the interaction weakens, though the previously identified hot spots are still present to an extent of about 20%. To inspect the interacting regions and the binding orientations in more detail, we generate contact map histograms. Such histograms for Cc and CcP at an ionic strength of 0 mM are depicted in Figure 6. Comparing the contact map histogram of 0 mM with the contact map of the crystal structure, it can be seen that the contacts of the crystal structure can be well-reproduced (Figure 6A,B). Some additional contacts can be identified, which arise from the analysis of the whole ensemble of complex formations. These contacts are located on the surface in the vicinity of main interacting residues. When key residues are engaged in contacts stemming from many different orientations, neighboring residues also get involved to a certain extent. These more...
distributed interactions underline the transient character of the encounter complex. Interestingly, for the key residues D148 and K149 of the secondary binding site, contacts with a population up to 11–25% can also be identified at low ionic strength. At 100 mM, the total docking events decrease, whereas the relative specificity of individual amino acids can still be sustained, as is expected with increasing ionic strength (Figure 6C). However, subtracting the normalized contact map histograms at 0 and 100 mM from each other allows identification of differences between the ensembles at different ionic strength values and highlights which residues are less involved in contacts at higher salt concentrations (Figure 6D). With the combination of contact map histograms and energy distribution plots, we can show that the experimentally characterized loss of the lowly populated binding sites and the shift of the contact surface induced by a higher salt concentration can be reproduced by our MC simulations. The various analysis and plotting methods provide a powerful tool to connect experimental data with structural information derived from the MC simulation.

**Influence of Cc Heme Oxidation States on the Formation of the Complex.** An important parameter that influences the association of electron-transfer proteins is their oxidation state. Experimentally, such an influence is often hard to access because many experimental methods are only applicable in certain redox states. Simulations are therefore a good alternative for analyzing the influence of this parameter. In MCMap, different redox states are modeled by assigning appropriate partial charges to the redox centers. For CcP, reduced Cc (Fe2+) state is the natural electron donor. In order to guarantee a high turnover, it would be biologically advantageous if oxidized Cc (Fe3+) would not bind as tightly. To probe this possibility, we simulated the association of CcP with both oxidized and reduced Cc at an ionic strength of 100 mM.

At first glance, the density distributions obtained for the two oxidation states of Cc are quite similar. However, the differences can be seen when the encounter density obtained for oxidized Cc is subtracted from the density obtained for reduced Cc (Figure 7A,B). The binding site of reduced Cc is more distributed, whereas the oxidized form with its more positive charge is more focused on the main interaction patch (Figure 7C,D).

While the difference map gives a reasonable approximation of the extent of the encounter complexes, the surface mapping tool PyCoALA grants a more detailed overview of which residues are mainly contributing to the complex formation. Utilizing PyCoALA, the difference between contact map histograms for the different oxidation states can be plotted and mapped to the protein surface. The residues that show a difference are depicted in Figure 8. It becomes clear that many interacting residues on CcP are the same for both oxidized and reduced states of Cc (shown in purple); however, the contacting areas of Cc shift slightly. With the interaction difference mapped on the surface, two distinct separated regions on Cc can be identified. Whereas K87 and K89 exclusively show many interactions with the peroxidase for reduced Cc, the interaction profile shifts to a region located around R13 and G83 for the oxidized Cc. Even if this change in the binding area is not dramatic, a difference between the oxidized and reduced ensembles can be seen, which may also influence the binding affinity.

**Ternary Complexes.** For the reduction of hydrogen peroxide to water, two electrons are required. Because CcP may bind several copies of Cc, two mechanistic scenarios are possible. Either, two reduced molecules of Cc bind consecutively; that is, the first Cc dissociates after the first electron transfer, giving way to a second Cc as an electron donor; or, two reduced Cc molecules bind to CcP at the same time and the two electron transfers occur without the necessity for dissociation.

We use our MC docking method to test if two copies of Cc could bind simultaneously to CcP. For this purpose, we use the crystal structure of the complex between CcP and one Cc as the receptor for two different MC simulations. In one simulation, we assume that the bound Cc is reduced, and in the other simulation, it is oxidized. The calculated electrostatic potential of the receptor represents thus the complex of CcP and the first bound Cc molecule. The ligand, that is, Cc that binds to the binary complex of Cc and CcP is always assumed to be reduced. For the simulation in which the bound Cc is reduced (Figure 9A), we obtained two binding spots: a small one near the interface between the bound Cc and CcP and a more extended one at the position of the secondary binding site. In the simulation with the bound oxidized Cc, the small binding spot instead virtually disappears (Figure 9B). In order to visualize the difference between the simulations, the density for the oxidized state was subtracted from the density for the reduced state (Figure 9C). The difference density
confirms the lower population of the binding spot near the interface but also shows that the outer part of the secondary binding site has a higher population in the reduced state. It seems that if a Cc molecule is already bound to the main binding site of CcP, the secondary binding site is highly promoted for interactions. The small binding spot near the interface can be explained by the influence of the remaining potential of the main binding site of CcP. The potential of CcP is not completely neutralized by the bound Cc and attracts a second Cc molecule. The changed binding behavior of the ligand to the different redox states of the CcP-bound Cc can also be seen in the population diagrams, where the energies of the different orientations are related to the distances between ligand heme and CcP heme (Figure 9D) and their probability of occurrence. The distributions look similar for both oxidation states with the energetically favored encounters being higher populated and more widely spread in the reduced form. Apparently, the oxidized form of Cc repels another Cc molecule more strongly, which can be seen by less low-energy encounters and a shift of the distribution to a higher heme-to-heme distance. Considering the plot of the distances between the ligand heme and CcP heme and the one between ligand heme and bound Cc heme (Figure 9E), we can clearly identify two clusters being predominant for the reduced simulation. Upon oxidation of the bound Cc, a shift of the hot spots from a short distance to a widely spread cluster equally distant to Cc and CcP heme can be observed. These findings agree with the idea that two binding Cc molecules repel each other electrostatically.46,52

Our simulation leads to a view that the oxidation state in the ternary complex influences the binding behavior of a second Cc molecule. From our analysis, we would support a model in which two Cc molecules bind at the same time. The binding is electrostatically favored and even promotes the binding of Cc to the secondary binding site, while for a consecutive binding, the interactions are slightly destabilized, as long as an oxidized Cc molecule is still bound to CcP.

**CONCLUSIONS**

In this paper, we introduced the software suite MCMap and demonstrated the usage of the software for mapping protein–protein interaction of transient protein complexes highly depending on electrostatic interactions. The software uses a rigid-body docking approach and is particularly well-suited for analyzing transient protein complexes, for instance, involved in electron-transfer processes. In MCMap, an MC method with a ligand moving randomly in the electrostatic field of a receptor is applied. Acceptance of MC moves and thereby weighting of the electrostatic interaction energies is done according to the Metropolis MC criterion, resulting in a Boltzmann distribution of states. This approach allows for the identification of orientations that are energetically not favored but still have a certain likelihood of occurrence. Recent studies showed that also relatively high energy intermediates contribute to the successful formation of a complex and that the complex formation follows an energy path in an energy landscape containing specific and nonspecific complexes.53 The obtained ensemble covers a distribution of complexes which are characteristic for transient protein complexes. In this way, MCMap covers not only the surface regions of a receptor but maps also its adjacent surroundings for interactions. The program comes with a variety of analysis options which we demonstrate in this paper. These analysis tools help to relate the simulation results to experimental data, which can then be explained on a structural level. In particular, with PyCoALA, we present an analysis method to rapidly link 2D contact histograms to 3D structures of the receptor and the ligand. This tool allows for analyzing predicted protein–protein interaction hot spots for their structural properties. In the context of growing efforts to find drugs modulating interactions between proteins, an application of such an analysis strategy in a drug design context is conceivable. Furthermore, the possibility to map the difference between two related simulations to the protein surfaces provides insight in microscopic factors contributing to an altered system behavior. Such an approach may find application not only in the context of studying the impact of redox states on the binding behavior. It may also provide more insights on the effect of post-translational modifications or mutations on binding ensembles, for instance, in gene regulation. In addition to the analysis demonstrated herein of protein–protein interaction, MCMap is well-suited to also analyze heavily charged molecules, such as DNA or RNA. MCMap also opens the possibility to linearly scale up the mapping simulations. In the way, MCMap saves orientations, and all MC runs are independent of each other. By this means, MCMap can be executed multiple times in parallel with the same parameters but different starting seeds in order to get one big encounter ensemble. In the application to the
complex of CcP and Cc, we demonstrate how, for instance, the ionic strength of the solution or oxidation states of proteins can influence the binding. Moreover, our simulations suggest that a second Cc molecule can be repelled by a bound Cc in the ternary complex. These results give insights into how the complex formation is microscopically achieved and how theoretical modeling can be used as a computational microscope for a better understanding of experimental data.54

■ MATERIALS AND METHODS

Preparation of the Protein Models. The structure of the yeast complex Cc:CcP (PDB code: 2PCC, chain A and B)49 was used in our analysis. We used the fitting method CHELPG55 within ORCA56 (B3LYP, def2-TZVP basis set) to derive partial charges for cofactor, which were not part of the CHARMM27 force field.57 For Cc, we derived charges for a c-type heme, typically linked to the protein by two cysteines and coordinated by histidine and methionine in both states, reduced (Fe2+) and oxidized (Fe3+).

We added hydrogens and energy-minimized their position using CHARMM16 with steepest descent and conjugate gradient methods in cycles of 500 steps each. During minimization, heavy atoms were kept fixed. The protonation states of all amino acids were determined with titration studies using an in-house modified version of MEAD58 and GMCT.59 The residues E11, D224, D235, and E267 in CcP were found to be protonated at pH = 7 and were treated as such in further analysis.

Calculation of the Electrostatic Potentials. Electrostatic potentials were calculated using APBS.37 The ionic strength was set to 0, 10, 20, 50, and 100 mM. Relative permittivity constants of 4 and 80 were used for the protein and the solvent, respectively. For the MC simulation, the electrostatic potential was represented on cubic grids with 225 grid points in each direction and grid spacings of 2.0 and 1.0 Å for the coarse and the fine grid, respectively. The temperature was set at 300 K. Radii of 1.4 and 2.0 Å were used for the solvent and the ions, respectively.

Simulation of Association Using MCMap. For protein–protein docking, we used MCMap, our software for performing MC docking simulations presented herein. We performed 10^8 runs with a total of 10^5 MC steps each for 0, 10, 20, 50, and 100

Figure 9. Distribution of configurations for the ternary complex consisting of the crystal complex CcP–Cc (PDB: 2PCC) and an additional Cc molecule as a ligand. In one simulation, the bound Cc is oxidized (A), and in the other, it is reduced (B). The isosurfaces represent half the maximal density of all orientations. The difference between the ensembles (Fe^{2+}–Fe^{3+}) is shown in (C), with the positive and the negative isosurfaces colored blue and red, respectively. For each simulation, the shortest heme-to-heme distance for the ligand to the CcP is plotted vs the electrostatic interaction energy (D). Another way to identify orientational changes is to analyze the population of the various heme-to-heme distances from the ligand Cc to the complex Cc as a function of the heme-to-heme distance of the ligand Cc to CcP. In all distance mappings, the ensemble distribution seems to have a clear edge at the top border. This edge is caused by the maximal separation of the ligand and the receptor considered in the analysis and is not a feature of the encounter complex (E).
mM ionic strength. The simulations were initiated with a center of mass separation of $R_{\text{in}} = 130 \, \text{Å}$ between the receptor and the ligand. A maximum center of mass separation of $R_{\text{out}} = 180 \, \text{Å}$ was allowed. The MC run was reset after 50 consecutively rejected MC steps. The temperature was set to 300 K. The electrostatic potential maps with grid spacings of 1 and 2 Å were used as inner and outer potential grids, respectively. A maximal displacement of 3 Å and a maximal rotation of 5 rad were used. Calculations using OpenDX files (such as subtractions of densities) were done using the tool DXMATH, which is a part of APBS.37

**Analysis of the Simulation Results.** For analyzing all orientations of a simulation ensemble, print-coor was used. For generating the contact map histograms, a distance of 8.0 Å was chosen. The heme-to-heme distances were calculated with the subprogram min-dist, which determines the minimal distance between two given structures or structural elements. In order to compare MCMAP runs with a different number of orientations in the encounter ensemble, the histograms were normalized to the total number of orientations. In this way, the histograms can be subtracted from each other to identify differences. For a better graphical representation, the histograms were scaled by the inverse of the highest absolute value of the histogram to lead to a maximal value of one for all histograms. For visualizing residues with major differences in the contacts between two simulations, PyCoALA and PyMOL were used.42

**AUTHOR INFORMATION**

**Corresponding Author**
*E-mail: matthias.ullmann@uni-bayreuth.de. Phone: +49 (0) 921 553545. Fax: +49 (0)921 553071.*

**ORCID**
Johannes M. Foerster: 0000-0003-0442-9413
G. Matthias Ullmann: 0000-0002-6350-798X

**Present Address**
1Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS gGmbH), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany (I.P.).

**Author Contributions**
‡J.M.F. and I.P. contributed equally to the work and are considered as co-first authors.

**Notes**
The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**
This work was supported by the German Science Foundation (DFG; GRK1640).

**REFERENCES**

(1) Kostić, N. M. Diprotein Complexes and Their Electron Transfer Reactions. In *Metal Ions in Biological Systems: Volume 27*; Sigel, H., Sigel, A., Eds.; CRC Press, 1991; pp 97–127.

(2) Kostić, N. M. Dynamic Aspects of Electron-Transfer Reactions in Metalloligographic Complexes. In *Metal-Containing Polymeric Materials*; Pittman, C. U., Carragher, C. E., Zeldin, M., Sheats, J. E., Culbertson, B. M., Eds.; Springer US: Boston, MA, 1996; pp 491–500.

(3) Hart, S. E.; Schlub-Ridley, B. G.; Delon, C.; Bendall, D. S.; Howe, C. J. Role of Electrostatic in the Interaction between Cytochrome f and Plastocyanin of the Cyanobacterium *Phormidium laminosum*. *Biochemistry* 2002, 41, 3279–3285.

(5) Scaru, S.; Foerster, J. M.; Timmer, M.; Ullmann, G. M.; Ubbink, M. Loss of electrostatic interactions causes increase of dynamics within the plastocyanin-cytochrome f complex. *Biochemistry* 2013, 52, 6615–6626.

(6) van Zundert, G. C.; Hudson, B. M.; Keedy, D. A.; Fonseca, R.; Heller, A.; Suresh, P.; Burrell, K.; Day, T.; Fraser, J. S.; van den Bedem, H. qFit-ligand reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps. *2018. bioRxiv 253439.*

(7) Qin, J.; Gronenborn, A. M. Weak protein complexes: challenging to study but essential for life. *FEBS J.* 2014, 281, 1948–1949.

(8) Müller, I. Guidelines for the successful generation of protein–ligand complex crystals. *Acta Crystallogr., Sect. D: Struct. Biol.* 2017, 73, 79–92.

(9) Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. Replica exchange simulations of transient encounter complexes in protein–protein association. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 12855–12860.

(10) Ubbink, M. The courtship of proteins: Understanding the encounter complex. *FEBS Lett.* 2009, 583, 1060–1066.

(11) Northrup, S. H.; Erickson, H. P. Kinetics of Protein-Protein Association Explained by Brownian Dynamics Computer Simulation. *Proc. Natl. Acad. Sci. U.S.A.* 1992, 89, 3338–3342.

(12) Grove, T. Z.; Ullmann, G. M.; Kostić, N. M. Simultaneous true, gated, and coupled electron-transfer reactions and energetics of protein rearrangement. *J. Inorg. Biochem.* 2012, 106, 143–150.

(13) Scaru, S.; Foerster, J. M.; Ullmann, G. M.; Ubbink, M. Role of hydrophobic interactions in the encounter complex formation of the plastocyanin and cytochrome f complex revealed by paramagnetic NMR spectroscopy. *J. Am. Chem. Soc.* 2013, 135, 7681–7692.

(14) Perkins, J. R.; Diboun, I.; Dessailly, B. H.; Lees, J. G.; Oreno, C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure* 2010, 18, 1233–1243.

(15) Di Nola, A.; Roccatano, D.; Berendsen, H. J. C. Molecular dynamics simulation of the docking of substrates to proteins. *Proteins: Struct., Funct., Genet.* 1994, 19, 174–182.

(16) Alonso, H.; Bliznyuk, A. A.; Greedy, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* 2006, 26, 531–568.

(17) Pflemdtner, J.; Volkman, N.; Hanein, D.; Dalhaimer, P.; Pollard, T. D.; Voth, G. A. Key Structural Features of the Actin Filament Arp2/3 Complex Branch Junction Revealed by Molecular Simulation. *J. Mol. Biol.* 2012, 416, 148–161.

(18) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* 2015, 31, 64–74.

(19) Gάbdoulline, R. R.; Wade, R. C. On the Contributions of Diffusion and Thermal Activation to Electron Transfer between *Phormidium laminosum* Plastocyanin and Cytochrome f : Brownian Dynamics Simulations with Explicit Modeling of Nonpolar Desolvation Interactions and Electron Transfer Events. *J. Am. Chem. Soc.* 2009, 131, 9230–9238.

(20) Gάbdoulline, R. R.; Wade, R. C. Brownian Dynamics Simulation of Protein-Protein Diffusional Encounter. *Methods 1998*, 14, 329–341.

(21) Gάbdoulline, R. R.; Wade, R. C. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol.* 2001, 306, 1139–1155.

(22) Ullmann, G. M.; Knapp, E.-W.; Kostić, N. M. Computational Simulation and Analysis of Dynamic Association between Plastocyanin and Cytochrome f Complex: Brownian Dynamics Simulations with Explicit Modeling of Nonpolar Desolvation. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107, 26, 531–568.

(23) Liu, M.; Wang, S. MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. *J. Comput-Aided Mol. Des.* 1999, 13, 435–451.

(24) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel
applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* 1996, 10, 293–304.

(25) Zhang, Z.; Schindler, C. E. M.; Lange, O. F.; Zacharias, M. Application of Enhanced Sampling Monte Carlo Methods for High-Resolution Protein-Protein Docking in Rosetta. *PLoS One* 2015, 10, e0125941.

(26) Scanu, S.; Förster, J.; Finiguerra, M. G.; Shabestari, M. H.; Huber, M.; Ubbink, M. The complex of cytochrome c and plastocyanin from Nostoc sp. PCC 7119 is highly dynamic. *ChemBioChem* 2012, 13, 1312–1318.

(27) Díaz-Moreno, I.; Hulsker, R.; Skubak, P.; Förster, J. M.; Cavazzini, D.; Finiguerra, M. G.; Díaz-Quintana, A.; Moreno-Beltrán, B.; Rossi, G. L.; Ullmann, G. M.; Pannu, N. S.; De La Rosa, M. A.; Ubbink, M. The dynamic complex of cytochrome c6 and cytochrome f studied with paramagnetic NMR spectroscopy. *Biochim. Biophys. Acta Bioenerg.* 2014, 1837, 1305–1315.

(28) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. Visualization of the Encounter Ensemble of the Transient Electron Transfer Complex of Cytochrome c and Cytochrome c Peroxidase. *J. Am. Chem. Soc.* 2010, 132, 241–247.

(29) Volkov, A. N.; Worral, J. A. R.; Holtzmann, E.; Ubbink, M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 18945–18950.

(30) Iwahara, J.; Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* 2006, 440, 1227–1230.

(31) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* 2006, 444, 383–386.

(32) Metropolis, N.; Ulam, S. The Monte Carlo method. *J. Am. Stat. Assoc.* 1949, 44, 335–341.

(33) Altschul, A. M.; Abrams, R.; Hogness, T. R. Cytochrome c Peroxidase. *J. Biol. Chem.* 1940, 136, 777–794.

(34) Beetlestone, J. The Oxidation of Cytochrome c by Cytochrome c Peroxidase. *Arch. Biochem. Biophys.* 1960, 89, 35–40.

(35) Xie, Z.-R.; Chen, J.; Wu, Y. Predicting Protein-Protein Association Rates using Coarse-grained Simulation and Machine Learning. *Sci. Rep.* 2017, 7, 46622.

(36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1983, 4, 187–217.

(37) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* 2001, 98, 10037–10041.

(38) Ullmann, G. M.; Bombara, E. Continuum Electrostatic Analysis of Proteins. In *Protein Modelling*; Náray-Szabó, G., Ed.; Springer International Publishing: Cham, 2014; pp 135–163.

(39) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2012, 2, 73–78.

(40) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules. An experience report on the MEAD project. In *Scientific Computing in Object-Oriented Parallel Environments: First International Conference, ISCOPE 97 Marina del Rey, California, USA, December 8–11, 1997 Proceedings*; Ishikawa, Y., Odlehoef, R. R., Reynolds, J. V. W., Tholburn, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1997; pp 233–240.

(41) Metropolis, N.; Ulam, S. The Monte Carlo method. *J. Comput. Chem.* 1990, 11, 361–373.

(42) Metropolis, N.; Ulam, S. The Monte Carlo method. *J. Comput. Chem.* 1990, 11, 361–373.

(43) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M.; All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 1998, 102, 3586–3616.

(44) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules. An experience report on the MEAD project. In *Scientific Computing in Object-Oriented Parallel Environments: First International Conference, ISCOPE 97 Marina del Rey, California, USA, December 8–11, 1997 Proceedings*; Ishikawa, Y., Odlehoef, R. R., Reynolds, J. V. W., Tholburn, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1997; pp 233–240.

(45) Grassie, H.; Wang, K.; McKee, S.; Wang, X.; Waldner, J. L.; Pelack, G. J.; Durham, B.; Millet, F. Control of formation and dissociation of the high-affinity complex between cytochrome c and cytochrome c peroxidase by ionic strength and the low-affinity binding site. *Biochemistry* 1996, 35, 15800–15806.

(46) Nakane, S.; Vitello, L. B.; Erman, J. E. Characterization of Four Covalently-Linked Yeast Cytochrome c/Cytochrome c Peroxidase Complexes: Evidence for Electrostatic Interaction between Bound Cytochrome c Molecules. *Biochemistry* 2006, 45, 14371–14378.

(47) Zhou, J.; Huber, M.; Ubbink, M.; Van Nuland, N. A. M. Mapping the encounter state of a transient protein complex by PRE NMR spectroscopy. *J. Biomol. NMR* 2010, 48, 225–236.

(48) Zhou, J.; Hoffmann, B. Stern-volmer in reverse: 2:1 stoichiometry of the cytochrome c-cytochrome c peroxidase electron-transfer complex. *Science* 1999, 265, 1693–1696.

(49) Pelletier, H.; Kraut, J. Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c. *Science* 1992, 258, 1748–1755.

(50) Page, T. R.; Hoffmann, B. M. Control of Cyclic Photoinitiated Electron Transfer between Cytochrome c Peroxidase (W191F) and Cytochrome c by Formation of Dynamic Binary and Ternary Complexes. *Biochemistry* 2015, 54, 1188–1197.

(51) Ponder, W. A. D.; Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules. An experience report on the MEAD project. In *Scientific Computing in Object-Oriented Parallel Environments: First International Conference, ISCOPE 97 Marina del Rey, California, USA, December 8–11, 1997 Proceedings*; Ishikawa, Y., Odlehoef, R. R., Reynolds, J. V. W., Tholburn, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1997; pp 233–240.

(52) Ullmann, R. T.; Ullmann, G. M. GMCT: A Monte Carlo simulation package for macromolecular receptors. *J. Comput. Chem.* 2006, 37, 6465–6475.