Tracing entire operation cycles of molecular motor hepatitis C virus helicase in structurally resolved dynamical simulations

Holger Flechsig1 and Alexander S. Mikhailov

Department of Physical Chemistry, Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, 14195 Berlin, Germany

Edited* by Gerhard Ertl, Fritz Haber Institute, Berlin, Germany, and approved October 20, 2010 (received for review September 29, 2010)

Hepatitis C virus helicase is a molecular motor that splits duplex DNA while actively moving over it. An approximate coarse-grained dynamical description of this protein, including its interactions with DNA and ATP, is constructed. Using such a mechanical model, entire operation cycles of an important protein machine could be followed in structurally resolved dynamical simulations. Ratcheting inchworm translocation and spring-loaded DNA unwinding, suggested by experimental data, were reproduced. Thus, feasibility of coarse-grained simulations, bridging a gap between full molecular dynamics and reduced phenomenological theories of molecular motors, has been demonstrated.

Results

Fig. 1 shows the ribbon representation of HCV helicase together with the constructed elastic network of this protein. Motor domains I and II translocate along the DNA strand, indicated in the ribbon structure. The ATP binding pocket is located on the right side of the motor domain I. On the left side of domain II, an important atomic group is present: Only when this atomic group comes into contact with ATP is the hydrolysis reaction possible (19). Thus, to allow ATP hydrolysis, conformational motions bringing together domains I and II are needed. In the elastic network, each (point-like) particle corresponds to a residue. Two particles are connected by an elastic string if the distance between them is less than a predefined value. Two elastic chains of identical beads; two strands were bridged by deformable links that could get broken if strong enough forces were applied. Links between DNA and motor domains of HCV helicase could be established and removed, depending on the configuration dynamics. Chemical details of ATP binding and hydrolysis were neglected in this coarse-grained description. The ATP was roughly modeled as one (fictitious) ligand particle that was placed into the actual ATP binding pocket, forming several stretched elastic links with the neighboring residues. The hydrolysis reaction and product release were approximately taken into account, assuming that a change in the ligand particle occurred, leading to the disappearance of all links connecting it to the protein.

With such simplifications, we could trace in numerical simulations the consequent cycles of the molecular motor and observe its active translocation along the DNA strand, leading to DNA unzipping.

Analyzing conformational relaxation processes in HCV helicase, we have found that, like other motor proteins (10, 12), it is able to perform well-defined conformational motions, robust against perturbations. Such motions are functional: They can bring into contact two motor domains of HCV helicase, thus making ATP hydrolysis possible. Additionally, they lead to conformational changes in the DNA binding cleft that are essential for translocation.

To simulate complete operation cycles of HCV helicase, interactions with DNA and ATP had to be further included into the dynamical description. DNA strands were modeled by us as elastic chains of identical beads; two strands were bridged by deformable links that could get broken if strong enough forces were applied. Links between DNA and motor domains of HCV helicase could be established and removed, depending on the configuration dynamics. Chemical details of ATP binding and hydrolysis were neglected in this coarse-grained description. The ATP was roughly modeled as one (fictitious) ligand particle that was placed into the actual ATP binding pocket, forming several stretched elastic links with the neighboring residues. The hydrolysis reaction and product release were approximately taken into account, assuming that a change in the ligand particle occurred, leading to the disappearance of all links connecting it to the protein.

With such simplifications, we could trace in numerical simulations the consequent cycles of the molecular motor and observe its active translocation along the DNA strand, leading to DNA unzipping.

Author contributions: A.S.M. designed research; H.F. performed research; H.F. and A.S.M. analyzed data; and H.F. and A.S.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.
†To whom correspondence should be addressed. E-mail: flechsig@hi-berlin.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014631107/-/DCSupplemental.

References

1. F. Haber Institute, Berlin, Germany, and approved October 20, 2010 (received for review September 29, 2010).
localization in the ATP binding pocket (cf. Fig. 1A and SI Text). One can clearly see that local perturbations in this region are inducing the same generic, well-defined conformational motions.

When ATP arrives, its interactions with the protein lead to the appearance of some local deformations in the ATP pocket region. In the framework of the normal mode analysis, it has been previously shown that, by applying an experimentally recorded pattern of local ATP-induced deformations (taken, however, from a different protein, bovine F1-ATPase), conformational changes, which are similar to those described above, become induced (25). In the coarse-grained elastic-network model with residues replaced by single particles, chemical structure of ATP and details of its interactions with the residues cannot be fully accounted. Our above results suggest that such details are not actually important when ligand-induced mechanochemical motions are considered.

With this observation, we proceeded further in the dynamical description. To remain at the coarse-grained description level, ligand binding was imitated in our model by placing an additional substrate particle into the ATP binding pocket. The new particle was connected by four links to the neighboring network nodes (see Fig. S1). These elastic links were stretched (i.e., their length was longer than the natural length), and thus energy was supplied into the system. Initially localized only in the four new links, deformations spread over the network and induced motor domain motions. These motions represented conformational relaxation processes of the network–ligand complex. They were finished when the equilibrium state of the complex was reached.

Fig. 2B displays the trajectory P → Q of the conformational motion induced by ligand binding. In Fig. 2C, conformational transition induced by ligand binding is shown. The most important effect is that domain II moves toward the motor domain I, so that the gap between them gets closed.

A reaction converting ATP into ADP and phosphate should take place when the gap is closed and atomic groups on the left side of domain II come into contact with ATP (16). After that, the products should be released. These chemical processes cannot be resolved in the considered coarse-grained approach. However, we could still roughly imitate them in our model. To do this, we assumed that, as the equilibrium state of the network–ligand complex is approached, the nature of the ligand becomes changed. It gets converted from the substrate to the product particle. In contrast to the substrate, the product particle has no interactions with the protein. Thus, the links connecting the ligand to the network disappear once the conversion has taken place, and the product is released. After the release of the product particle, the free network finds itself in a conformation different from its true equilibrium state. Hence, the reverse conformational relaxation process should begin. In this process, corresponding to the trajectory Q → P in Fig. 2B, the motor domain II moves back away from the domain I, restoring the initial configuration (the cyclic mechanochemical motions can be seen in Movie S1).

It is known that, as a result of ATP binding and hydrolysis, two motor domains translocate along the single DNA strand. To describe such a process, interactions between the protein and the DNA strand had to be included into the model. The DNA strand was modeled as a semiflexible polymer chain with identical strands localized in the ATP binding pocket (cf. Fig. L4 and SI Text). One can clearly see that local perturbations in this region are inducing the same generic, well-defined conformational motions.

Comparing protein conformations in its equilibrium ligand-free and ligand-bound states (Fig. 2C), a significant difference, best seen in the side view in Fig. 2C, can be noticed. The two motor domains have roughly the shapes of claws that can grasp the DNA strand. In the equilibrium ligand-free state P, the claw of the left domain I is moved down toward the DNA,
whereas the claw of the right domain II is lifted. In contrast to this, in the equilibrium ligand-bound state Q the left claw is lifted and the right claw is moved down (Fig. 2C). Thus, the DNA strand may be held by the left domain I without the ligand and by the right domain II when the ligand is present.

Interactions between DNA and the protein are complicated, and structural details of the DNA and the protein are important in order to determine them accurately. This lies, however, beyond the resolution level of the current coarse-grained description. In our rough model, we could only phenomenologically imitate such interactions. To do this, we assumed that elastic links could be established between specific particles (residues) in the cleft regions of two motor domains and the beads (nucleotides) forming the polymer chain. Such additional links were established or disappeared depending on the ligand presence (see Materials and Methods).

With this model extension, the translocation cycle of HCV helicase could be reproduced in our simulations (see Fig. 3 and Movie 2). In the equilibrium ligand-free state, the DNA is attached by a link to the left domain I. Ligand binding induces a quick lift of the claw formed by domain I, whereas the link attaching DNA to domain I disappears. At the same time, the claw of domain II moves down and a link attaching DNA to this domain becomes formed. After that, slower relative motion of domain II toward domain I, while grasping the DNA strand, takes place. When the new equilibrium (of the ligand–network complex) is reached, the ligand disappears. Now, the grip is changed. The left domain I moves down to DNA and a link attaching it to one of the polymer beads is established. The right domain moves up and the link between it and the DNA strand disappears. Then, domain II slowly returns to its position, but without carrying with it the DNA strand. After each ligand-binding and ligand-release cycle, the protein returns to its initial conformation, but it becomes translocated by one nucleotide along the DNA strand.

The model was further extended to incorporate DNA unzipping (Fig. 4 and Movie S3). Here, the lower DNA strand was modeled as a semiflexible polymer chain bridged by deformable links to the upper strand. Such links become broken, if stretched above a limit (see Materials and Methods). As the motor domains I and II repeatedly translocate over the upper strand, they pull domain III into the space between two DNA strands. Thus, the force tearing apart the two strands is generated. As it is built up, stress gets accumulated in the links bridging the strands and, eventually, one bridge gets broken, unzipping the DNA by a base pair.

Discussion

Simplified descriptions for the protein, the DNA, and for the interactions between them and with the ATP ligand have been used in our study. The protein was modeled as a network of identical particles connected by identical elastic links. DNA was also described as an elastic chain of identical beads. Interactions between the protein and DNA were phenomenologically incorporated into the model, assuming that elastic links connecting one of the network particles and a bead in the chain are established or broken. Finally, binding of ATP, hydrolysis, and release
of the products were only roughly accounted for, by assuming that a fictitious substrate ligand particle gets bound to the network and is converted into the product, which is then immediately released. Thermal fluctuations and hydrodynamical interactions were neglected here.

It is therefore remarkable that even such a simple model could reproduce, in a structurally resolved manner, principal operation of HCV helicase and allowed us to trace entire cycles of this molecular machine. In the simulations, inchworm translocation and spring-loaded DNA unwinding, previously proposed, based on the experimental data, as the operation mechanism of HCV helicase (19, 26), have been reproduced. Our results are in agreement with the experiments by Gu and Rice (27), where conformational snapshots along the cycle of HCV helicase could be obtained.

HCV helicase, as a virus protein, is subject to strong genetic variability. Proofreading and error correction mechanisms are not employed by viruses and, generally, all copies of their proteins are different. Despite strong variations, the function of a protein must be maintained if this is crucial for virus reproduction. Extreme functional robustness should be characteristic for the molecular motors of viruses.

Currently, two principal operation mechanisms are being discussed for protein motors. They can either work similarly to macroscopic mechanical machines, performing ordered activated internal motions, or work as Brownian ratchets and employ fluctuations. Stochastic ratchets can be energetically more efficient and, if the efficient use of energy is of crucial importance, they would probably be preferred by Nature. Indeed, there are experimental indications that myosin motors of a cell act as Brownian ratchets (30). For viruses, energetic efficiency should not, however, play an important role. Instead, evolutionary pressure should be driving toward the development of virus motors with high robustness against structural perturbations.

In the investigations of HCV helicase, a number of conserved motifs, i.e., of the residues whose presence is essential for virus replication, have been previously identified (18, 19, 31). In our phenomenological modeling of interactions between DNA and the protein and in the description of ligand binding, these known conserved motifs were always chosen (see Materials and Methods). Furthermore, we have also performed a preliminary analysis (to be published separately) of structural robustness and could see that the operation of this molecular machine was not sensitive to local changes in its structure.

In the present study, our aim was to demonstrate that structurally resolved coarse-grained descriptions of the operation of molecular motors are feasible. Unwinding of DNA in bursts comprising several base pairs was experimentally observed (23, 26),
whereas unwinding by only one pair in each cycle has been seen in our simulations. It is, however, plausible that, by tuning the parameters, gradual accumulation of strains and subsequent breaking of a group of links can also be reproduced. Thermal conformational fluctuations and interactions with the solvent can be incorporated into the elastic-network description (see, e.g., ref. 32). As a coarse-grained approach, it cannot, however, resolve chemical details that are important to correctly model processes of ligand binding or dissociation and protein–DNA interactions. In the future, a combination of elastic-network modeling with full molecular dynamics simulations may need to be undertaken.

Our study has been focused on conformational aspects. Physical processes of ATP binding, its hydrolysis, and products release were not considered, and their effects were only imitated here. In our simulations of machine cycles, the next cycle started immediately once the previous cycle had been completed. Thermal fluctuations, allowing upfrill conformational motions, were not included into the description. Moreover, reverse processes of product binding, of its back conversion into the substrate, and of substrate dissociation were all neglected. Thus, the considered mechanism was irreversible; it corresponds to an idealized situation where substrate is supplied in abundance and products are instantaneously removed.

The overall direction of any catalytic reaction, including that of ATP hydrolysis by HCV helicase, is determined by difference in chemical potentials of its substrate and products. At equilibrium inside a closed volume, the chemical potentials are equal and directionality disappears. On the other hand, if the chemical potential of the products is maintained higher than that of the substrate, reverse catalytic events should prevail, so that, on the average, the same reaction would proceed in the opposite direction. The directionality issues in the operation of HCV helicase, involving mass action, should be a subject of separate investigations.

Materials and Methods

Construction of the Network. To construct the elastic network of HCV helicase we have used the apo structure of the protein (18) deposited in the RCSB Protein Data Bank (PDB ID code 1HIE). The PDB provides two independent sets, and we have chosen chain A because it yields more complete structural information (N = 442 residues). Using these data, the elastic network has been constructed by treating each amino acid (the entropic residue) as a single site represented by a point-like particle. Each particle is placed at the equilibrium position of the α-carbon atom in the main chain of the respective residue, denoted by $\mathbf{R}_i^0$ for particle $i$. Then, two particles are connected by an elastic spring if the distance between them is less than some cutoff distance $l_c$. The connection pattern of the network is stored in matrix $A$ with the elements $A_{ij} = 1$, if $d_{ij} - d_{ij}^0 < l_c$, and $A_{ij} = 0$, otherwise. As the cutoff distance, we have chosen $l_c = 8\,\AA$. The elastic energy for the network is $U = (k/2)\sum_i A_{ij} (d_{ij} - d_{ij}^0)^2$, where $k$ is the stiffness constant (the same for all links) and $d_{ij} = |\mathbf{R}_i - \mathbf{R}_j|$ is the distance between particles $i$ and $j$ in a deformed configuration with the particles at the positions $\mathbf{R}_i$. By definition, the equilibrium conformation of the network corresponds to the global energy minimum $U = 0$.

Relaxation Dynamics. Conformational dynamics of the protein is determined by a set of differential equations. For each particle, the equation of motion is $d\mathbf{R}_i/dt = -\nabla U/\nabla \mathbf{R}_i$. Here, $d\mathbf{R}_i/dt$ is the velocity of the particle $i$ at time $t$ and $\Gamma_i$ is its mobility, the same for all particles. These equations correspond to the overdamped limit valid at low Reynolds numbers with hydrodynamical interactions neglected. The system of dynamical equations reads as

$$
\frac{d\mathbf{R}_i}{dt} = -\sum_{j=1}^N A_{ij} \left( |\mathbf{R}_i - \mathbf{R}_j| - |\mathbf{R}_i^0 - \mathbf{R}_j^0| \right) \frac{\mathbf{R}_i - \mathbf{R}_j}{|\mathbf{R}_i - \mathbf{R}_j|}.
$$

These equations describe conformational relaxation as downhill motion in the elastic energy landscape. The Hooke forces acting on the particles depend linearly on deformations of the springs. However, because distances $d_{ij}$ represent nonlinear functions of spatial coordinates, $d_{ij} = |\mathbf{R}_i - \mathbf{R}_j| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$, Eqs. 1 are actually nonlinear. Here and below, dimensionless rescaled time $\tau t$ is used, which allows us to eliminate the spring constant $k$ and the particle mobility $\Gamma_i$ from the dynamical equations. To visualize conformational protein motions, we have tracked distance changes between the three labels $\Delta d_{ij}$, $\Delta d_{ij}$, and $\Delta d_{ij}$, with, e.g., $\Delta d_{ij} = d_{ij} - d_{ij}^0$. Relaxation trajectories are in general represented in a 3D space with normalized coordinates $u_{12}$, $u_{23}$, and $u_{13}$, with, e.g., $u_{12} = \Delta d_{12}/d_{12}^0$.

Modeling of Interactions with ATP and Its Hydrolysis. Interactions of proteins with energy-bringing ATP molecules rely on complex chemical processes that cannot be resolved in our purely mechanical model. In consistence with the coarse-grained descriptions of protein dynamics, we only roughly emulated interactions with ATP through the introduction and removal of an additional particle (a fictitious ligand). In our model, binding of an ATP molecule was mimicked by placing the ligand into the ATP binding region of domain $i$, in the vicinity of the conserved motifs Walker A and Walker B. Under binding, the ligand creates four new elastic links to the particles (residues) from the conserved motifs. Such links are initially deformed (stretched). The energy of the network–ligand complex is

$$
U_{\text{complex}} = c/2\sum_{i,j} A_{ij} (d_{ij} - d_{ij}^0)^2 + c/2 \sum_{i,j,i',j'} A_{ij} (d_{ij} - d_{ij}^0)^2.
$$

The first term is the elastic energy of the free network and the second term is the energy of ligand–network interactions. Here, $c_{1}, c_{2}$ are the parameters of the particles to which the ligand (index $N + 1$) is bound and $l_{\text{bat}}$ are their common natural lengths. To roughly take into account the hydrolysis reaction and products release, we assumed that, when the equilibrium state of the complex is reached, the links connecting the ligand to the network disappear and the ligand is thus released. Further details are provided in the SI Text.

Single DNA Strand. To demonstrate how ligand-induced cyclic motions of two motor domains can lead to translocation of HCV helicase along DNA, we have combined the elastic-network description for the protein with the semiflexible polymer modeling for the DNA. A single strand of DNA was modeled as a semiflexible chain of $M$ monomers (beads), where each bead represented a nucleotide and was connected to its neighboring beads by flexible links (see Fig. S3). The total elastic energy of the chain representing the stiff sugar–phosphate backbone of DNA is

$$
U_{\text{DNA}} = k_{el}/2 \sum_{i=1}^{M-1} (|\mathbf{r}_{i+1} - \mathbf{r}_i| - \sigma)^2 + k_{\text{bend}} \sum_{i=1}^{M-1} \left(1 - \cos \theta_{ij} \right)^2.
$$

The first term, a harmonic bonding potential, takes into account elastic deformations of the links in the chain. Here $\mathbf{r}_i$ is the position vector of bead $i$, $\sigma$ is the native length of the links, and $k_{el}$ is the stiffness constant, representing the stiff sugar–phosphate backbone of DNA. The second term is the elastic bending energy of the chain, with $\theta_{ij}$ being the angle between two adjacent links in the chain and $k_{\text{bend}}$ being the bending stiffness constant. The angle is defined as

$$
\cos \theta_{ij} = \frac{(\mathbf{r}_{i+1} - \mathbf{r}_i) \cdot (\mathbf{r}_{i+1} - \mathbf{r}_i)}{|\mathbf{r}_{i+1} - \mathbf{r}_i|^2},
$$

The overdamped relaxation motion of the beads is described by differential equations $d\mathbf{r}_i/dt = -\partial U_{\text{DNA}}/\partial \mathbf{r}_i$. For simplicity, we assumed that beads of the DNA had the same mobility as the particles forming the network, $\Gamma = \Gamma_i$. In our simulations, numerical values $k_{el}/\sigma = 20$ and $k_{\text{bend}}/\sigma = 15\,\AA^2$ were used, so that the DNA chain was much stiffer than the elastic network of the protein. We have chosen $\sigma = 6.5\,\AA$.

Duplex DNA. The duplex DNA consists of two strands with additional links that bridge opposite beads (see Fig. S4). Its total energy is the sum of the energies of the two strands and of the interaction potential, $U_{\text{DNA}} = U_{\text{DNA}}^{\text{strand}} + U_{\text{DNA}}^{\text{link}}$. Here, superscripts 1 and 2 are used to distinguish between the two chains. The interaction potential is

$$
U_{\text{int}} = D \sum_{i=1}^{M} \sum_{j=1}^{M} \left[ \exp(-a\Delta_i) - 1 \right]^2 + \kappa_i \sum_{i=1}^{M} \left(\cos \phi_i^{(1)}\right)^2 + \left(\cos \phi_i^{(2)}\right)^2,
$$

The first term is a sum of the Morse potentials, where $\Delta_i = |\mathbf{r}_i^{(1)} - \mathbf{r}_i^{(2)}| - l_{\text{bat}}$ is the elongation of a bridge link $i$. The second term depends on relative orien-

Flechsig and Mikhailov
tation of the bridge links and prevents shear motion of the DNA strands with respect to each other, taking into account shear stiffness of the duplex DNA. Here, the angles are defined as
\[
\cos \varphi_i^{(1)} = \frac{(\vec{r}_{i+1}^{(1)} - \vec{r}_i^{(1)}) \cdot (\vec{r}_i^{(2)} - \vec{r}_{i+1}^{(1)})}{|\vec{r}_{i+1}^{(1)} - \vec{r}_i^{(1)}||\vec{r}_i^{(2)} - \vec{r}_{i+1}^{(1)}|}.
\]

The Morse potential becomes flat if the elongation of a bridge link is much larger than the characteristic length 1/\(a\). Thus, the force disappears and the link becomes effectively broken. This allows us to model DNA unzipping processes. In simulations, numerical values \(a = 4.5 \text{ Å}^{-1}\), \(D/\sigma = 6.67 \text{ Å}^2\), \(l_d = 20 \text{ Å}\), and \(x_0 = 0.05 \text{ Å}^2\) were used. To simplify calculations, we have introduced a cutoff at \(\Delta = 5.5 \text{ Å}\), assuming that bridge interactions disappear completely for larger separations between two DNA beads.

Interactions with DNA. To model translocation, the single DNA chain was placed into the binding cleft. With \(\sigma = 6.5 \text{ Å}\), the separation by three nucleotides in the DNA strand agreed with the distance between the residues Thr269 of domain I and Thr411 of domain II, which are known to make contacts with DNA (19). During the ligand-induced cycle, interactions between motor domains of the protein and the DNA backbone have been roughly taken into account by introducing elastic orientation-dependent interactions (links) between one of these two residues in the motor domains and a bead in the DNA chain. Such links could be established or broken, controlled by ligand binding or detachment. Further details are provided in SI Text.

Separation of Duplex DNA. As the two motor domains translocate themselves along the upper strand of DNA, domain III becomes dragged into the space between the two strands. DNA cannot penetrate into the protein, and therefore the lower DNA strand should provide a barrier for the motion of helicase domain III. Actual interactions between DNA and the protein are complicated. They cannot be resolved in our coarse-grained description, and we take them only phenomenologically into account, by introducing repulsive interactions between the protein and the DNA strand. Further details are provided in SI Text.

ACKNOWLEDGMENTS. The authors are grateful to R. D. Astumian for his comments on microscopic reversibility aspects and thank R. Kapral and Y. Togashi for valuable discussions.