Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Adv Theory Simul. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

Adv Theory Simul. 2019 May ; 2(5): . doi:10.1002/adts.201800191.

Molecular Mechanisms of DNA Replication and Repair Machinery: Insights from Microscopic Simulations

Christopher Maffeo, **Han-Yi Chou**, **Aleksei Aksimentiev***

Department of Physics, Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign,1110 W Green St, Urbana, IL 61801, USA

Abstract

Reproduction, the hallmark of biological activity, requires making an accurate copy of the genetic material to allow the progeny to inherit parental traits. In all living cells, the process of DNA replication is carried out by a concerted action of multiple protein species forming a loose protein-nucleic acid complex, the replisome. Proofreading and error correction generally accompany replication but also occur independently, safeguarding genetic information through all phases of the cell cycle. Advances in biochemical characterization of intracellular processes, proteomics and the advent of single-molecule biophysics have brought about a treasure trove of information awaiting to be assembled into an accurate mechanistic model of the DNA replication process. In this review, we describe recent efforts to model elements of DNA replication and repair processes using computer simulations, an approach that has gained immense popularity in many areas of molecular biophysics but has yet to become mainstream in the DNA metabolism community. We highlight the use of diverse computational methods to address specific problems of the fields and discuss unexplored possibilities that lie ahead for the computational approaches in these areas.

1 Introduction

A biological cell uses the chemical structure of long DNA molecules to encode all the information required to produce its functional copy. The units of genetic information, adenine (A), thymine (T), guanine (G) and cytosine (C) nucleotides, differ from one another by just a handful of atoms. Yet, these minute differences prescribe the complimentary A-T and C-G base pairing and provide the physical mechanism for gene transcription. According to the central dogma of molecular biology [1], the information encoded in the nucleotide sequence of DNA is first transcribed into an RNA molecule of complementary nucleotide sequence, which is then used by a molecular factory—a ribosome—to synthesize proteins that perform virtually all cellular functions. This orderly flow of genetic information can, however, be disrupted by various dramatic events, the most common one being cell proliferation, Fig. 1. Preceding cell division is genome replication, a process that involves a concerted action of many proteins that perform efficient duplication of the genomic information while minimizing and correcting mistakes. In addition to replication errors, various environmental factors, such as reactive oxygen species or UV light, can damage a cell's DNA, which is why cells invest resources into genome maintenance machinery that

* aksiment@illinois.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

continuously inspects genome integrity and repairs DNA damage. Sometimes, the damage is too extensive to be repaired locally and requires another DNA molecule of similar (homologous) sequences to serve as a blueprint for repair. A disturbance of the DNA replication or repair processes can lead to numerous health conditions [2, 3] and is one of the main causes of aging [4, 5].

Much of what we know about the process of DNA replication and repair comes from the decades of biochemical analysis of cellular processes [8, 9]. Experimental methods have been developed to purify proteins and nucleic acids and examine their function in isolation, i.e. *in vitro.* Because of these efforts and recent advances in proteomics analysis, we now know what proteins perform DNA replication and repair functions, how many of them are present in a cell, and how the performance of individual components depends on the environmental conditions; we also know the interaction map between the components of the machinery. Furthermore, the structure of proteins involved in key DNA metabolism processes is known, at least partially, at the atomic resolution. However, most of the information derived from biochemical analysis reports on the properties of an ensemble of many proteins, which is invaluable in characterizing the net outcome of the replication or repair processes but provides limited detailed information about the molecular mechanisms.

Over the past twenty years, several experimental approaches have been developed to probe biochemical processes at the level of single molecules. Optical [10] and magnetic [11] traps have been developed to measure forces exerted by individual DNA binding proteins as they move along DNA. Fluorescent labels were covalently linked to proteins and DNA to monitor nanometer displacement of the labeled entities by measuring a Förster Resonance Energy Transfer (FRET) signal [12]. Optical traps and fluorescence measurement were combined [13] to directly obtain the force-versus-distance characteristics of DNA processing motors, sometimes revealing unexpected behavior [14]. Advances in super-resolution microscopy [15] illuminated spacial organization of DNA processing machinery in living cells. However, the single-molecule approaches have their own limitations. Optical trap measurements are difficult to perform *in vivo* [16], single-molecule FRET measurements are confined to near-surface subvolumes, and super-resolution imaging typically required cell fixing. Another common limitation is time resolution, which usually does not permit characterization of processes happening faster than a millisecond. The principle difficulty in characterizing the process of DNA replication at the single-molecule level is the multitude of protein species involved and the multiaxial character of the process.

Computational methods can complement experimental biochemical and biophysical studies by providing a much more detailed view of the system, though typically at sub-millisecond timescales. Starting from a structural representation of the system of interest, physics-based simulation methods can propagate the starting configuration in time, revealing the sequence of microscopic events that give rise to a molecular process at the all-atom, coarse-grained or even quantum levels. In the best cases, such simulations can reveal the actual mechanism of a protein’s function, providing interpretations of the experimental measurements and sometimes generating insights into possible drug targets or sites for genetic engineering.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In this review, we summarize recent modeling and simulation work on the core topics of DNA metabolism—replication, repair, and recombination, Fig. 1C. Our selection of systems is based solely on their popularity among modelers and is neither intended to provide a representative overview of the evolutionary development of the field nor is presented in any particular historical order. Though it is a subject of great interest, we will not discuss in detail the process of transcription, in particular, the numerous computational studies that examined DNA interactions with a variety of transcription factors. Nor will we discuss the studies of proteins responsible for maintaining the structure of chromatin, including histones, topoisomerases, cohesin/condensins, etc. Finally, with few exceptions we will not review the many chemical reactions occurring in the processes of DNA replication, damage and repair. Interested readers are directed to recent reviews of these subjects [17–27].

We start our review with a brief description of the most common computational methods used to study DNA systems. Table 1 links the systems and methods by providing explicit references to the studies of specific systems performed using specific methods. Following that, we review the computational studies pertaining to DNA replication, repair and homologous recombination processes. Each of these subsections is organized to cover essential elements of the processes, highlighting select simulation results. At the end of the review, we briefly describe opportunities lying ahead for computational biophysics approaches in the DNA metabolism area.

2 Numerical Methods

Since the molecular dynamics (MD) method was described more than five decades ago [190], it became one of the most powerful numerical tools in many disciplines of science because of the rapid progression of the computer hardware, algorithms and MD codes. Using a petascale supercomputer system and a state-of-the-art MD code [191], it is now possible to simulate a large biomolecular system containing over 10^7 atoms for a total simulated time of approximately 1 microsecond [192]. Such large-scale MD simulations can provide invaluable insights into atomic-scale interactions that give rise to biological processes that may be impossible to observe experimentally. Another advantage of the MD method is its empirical parameterization. As a result, a carefully designed and performed MD simulation can be expected to reproduce experimental observables, generating insights about the process of interest. Finally, the MD method serves as a foundation for other computational approaches such as hybrid quantum mechanics / molecular mechanics, enhanced sampling techniques, free energy calculations, and even to phenomenological or stochastic modeling. In this section, we will briefly describe the spectrum of computational approaches that have been applied to study DNA metabolism.

2.1 All-atom MD

The all-atom MD method represents atoms as point particles that move according to classical Newtonian dynamics:

$$m\frac{d^2\vec{r}}{dt^2} = \sum\left(\vec{F}_{\text{Coulomb}} + \vec{F}_{\text{non-bonded}} + \vec{F}_{\text{bonded}}\right), \quad (1)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where the force terms in Eq. 1 come from full Coulomb electrostatic interactions, other nonbonded interactions (usually van der Waals forces) and bonded interactions, which may contain two-, three- and four-body terms (usually bonds, angles and dihedrals). The interactions between atoms are described using empirical functional forms. Presently, AMBER [193], CHARMM [194], OPLS [195] and GROMOS [196] are the parameter sets (force fields) most commonly used for biomolecular simulations. Complementing force field development, several MD codes were written to make efficient use of massively parallel supercomputers, including NAMD [191], GROMACS [197], AMBER [198] and CHARMM [199]. Conventionally, one performs an MD simulation while coupling the system of interest to a thermostat (heat reservoir) to realize different statistical ensembles [200–206]. Even though all-atom MD is a good tool to investigate the dynamics of a biomolecule near its free energy minimum, MD often fails to sample the whole free energy landscape adequately and efficiently. Such sampling can be accomplished using the enhanced sampling methods.

2.2 Coarse-grained simulation

Although the all-atom MD method can now be used to simulate fairly large biomolecular systems, the characteristic time scale of many biological processes is not reachable by a straightforward all-atom MD simulation, even when using a massively parallel supercomputer. Coarse-graining offers a means to bridge the timescale gap. The basic idea is to employ the same clas-sical mechanics description of a biomolecular system as in the all-atom MD method, but using particles that represents collections of atoms rather than individual atoms.

Although there are many ways to derive a coarse-grained model of a biomolecular system [207], most of them can be categorized as being either top-down or bottom-up. In the top-down approach, one constructs and parametrizes a coarse-grained model to reproduce the physical properties of a target system, often disregarding the information about the system's chemical structure. In a bottom-up approach, a fine, often atomic level, description is used to construct a coarser representation by matching a many-body potential of the mean force (PMF). Many methods have been proposed for PMF matching, including iterative Boltzmann inversion [208], inverse Monte Carlo [209, 210] and multiscale coarse-graining [211, 212]. Both iterative Boltzmann and inverse Monte Carlo approaches match the pair distribution functions obtained using the coarse-grained potential to the reference distribution iteratively. The multiscale coarse-graining approach aims to match the coarse-grained and all-atom forces on each coarse-grained site. In practice, one often combines top-down and bottom-up approaches to derive a reliable coarse-grained model of good transferability [213–215].

Coarse-grained approaches have been used to study conformational transitions in proteins involved in DNA replication, for instance the PCNA clamp [40, 152], the DnaA protein [155] and the SSB protein [156]. Another example is a study of the thermodynamic and structural properties of the Holliday junctions that play a central role in the recombination process [159]. In general, a coarse-grained model brings computational efficiency at the expense of accuracy, with the most common deficiency being a poorly defined simulation time scale [216].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.3 Hybrid Molecular Mechanics / Quantum Mechanics (QM/MM)

Biological processes that involve chemical reactions or charge transfer cannot be described using a classical force field alone as they require explicit treatment of quantum mechanical processes. However, full *ab initio* simulation are still computationally prohibitive and impractical for systems containing more than a few hundreds of atoms. The seminal paper ofWarshel [217] introduced a hybrid quantum mechanics / molecular mechanics (QM/MM) approach, which has now become a standard method to approach chemical reaction and charge transfer problems.

Although there are many variants of the QM/MM method, they all share the same principle. A system of interest is divided into several regions: the MM region treated using a classical force field, the QM region treated using either density functional theory (DFT) or other self-consistent-field method [218], and, possibly, a third QM-MM coupling region where different QM/MM variants have their own formulations. One popular approach to QM/MM, the subtractive scheme, requires computing the energy of the whole system using a classical force field, the energy of the QM region using a QM method, and subtracting out the doubly-counted part, which is the MM energy in the QM region. The advantage of this approach is its relatively straightforward implementation in existing QM and MD codes [219]. The disadvantage is that one has to evaluate the QM region using a classical force field which may not be possible. In the additive scheme, one computes the energy in the QM subsystem using a QM method, the energy of the MM subsystem using the MM method and of the QM/MM coupling region using a model-specific method. The advantage of this approach is that one only has to treat the QM and MM regions using the corresponding methods. The coupling between the QM and MM regions is usually evaluated at the level of electrostatic interactions, which are considered to be dominant. Detailed description of different QM/MM schemes can be found in recent reviews [218, 220]. The QM/MM methods are particularly suitable for computational studies of DNA polymerase catalysis and reaction pathways [54, 129, 130, 135, 137–140]. Even though the conventional QM/MM method is widely used to investigate enzymatic reactions, its application is limited to small sub-systems containing only a few hundreds of atoms. It is also quite challenging sample biomolecular configurations while using *ab initio* QM/MM [221].

The above QM/MM methods require *ab initio* calculation of the electron structure of atoms in the QM region, which makes the calculation computationally demanding. There is, however, an alternative approach—the empirical valence bond (EVB) method [222–224], which makes computational studies of chemical reactions considerably more tractable. The EVB method describes the Hamiltonian of a reacting system as a mixture of the diabatic states corresponding to the classical valence bond structures describing the reactant, the intermediate states and the products of the reaction [223]. The diagonal terms of the EVB Hamiltonian are energies of the reactant and product states of the reaction. The off-diagonal terms are less obvious and one uses the empirical approximation for reproducing the barrier height [222] or fitting to other quantities [225]. Interested readers can find a much more detailed description of the EVB method in Ref. [226]. The EVB method was used to study the synthesis of DNA during replication [55,120,122, 151] and base excision repair of DNA damage [123]. The EVB method is much less computationally demanding than the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conventional *ab initio* QM/MM method and is able to faithfully describe catalytic effects and enzyme–solvent interactions. However the accuracy of the method strongly depends on the reliability of the empirical interaction potential coupling the reactant and product states, which maybe difficult to obtain.

2.4 Free energy calculation

As free energy determines the likelihood of biomolecular conformations, the strength of ligand binding and the pathway and rate taken by a biochemical reaction, several approaches have been developed to compute it. One of the most popular method is the free energy perturbation (FEP) proposed by R. W. Zwanzig [227]. FEP relates the free energy difference ΔF between the initial, A, and the final, B, states with the ensemble average of their energy difference averaged over all configurations in the initial state according to

$$\Delta F = F_A - F_B = -k_B T \ln \left\langle \exp[\frac{E_A - E_B}{k_B T}] \right\rangle_A , \tag{2}$$

where E_A and E_B are the Hamiltonians corresponding to states A and B. Even though Eq. 2 is exact, using it for calculation of the free-energy difference by conventional all-atom MD is formidable because of the inadequate sampling when the two states are separated by a high energy barrier. Therefore, FEP is commonly used in tandem with enhanced sampling techniques discussed below (Section 2.5). If we perform a Taylor expansion of Eq. 2 up to the second order with the assumption that the deviation in energies between the two states is small, we can express Eq. 2 as

$$\Delta F = \left\langle V \right\rangle - \frac{1}{k_B T} \left(\left\langle V^2 \right\rangle_A - \langle V \rangle_A^2 \right), \tag{3}$$

where V is the (small) difference in energy between A and B. One can attempt to compute the free energy difference between the states with either the exact or approximate expression using either Monte Carlo or MD to sample molecular configurations [228]. The FEP method has been used to study the fidelity of DNA replication by computing the free-energy of mismatched structures in complex with protein [121] and for evaluation of repair pathways [123].

An alternative to FEP is the linear response approximation (LRA) [229–231]. The LRA method is based on the assumption that the free energy is a parabolic function of the reaction coordinate chosen to describe the transition between two states. In practice, LRA is applied to compute the electrostatic part of the free energy difference [232]. A related method, linear interaction energy (LIE), uses the approximate LRA for the electrostatic part and scales the average van der Waals interactions. LRA and LIE method are typically used to evaluate the binding energy of a protein–ligand complex, as reviewed in Refs. [232] and [233]. These methods can be viewed as approximations to the FEP method and are typically used to study biomolecular binding processes [232] and to reveal how biomolecular interactions guide

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

catalytic activity [46, 119, 120, 124, 125], but their accuracy is limited because of the underlying approximations.

2.5 Enhanced Sampling Techniques

One of the most popular sampling approaches is the umbrella sampling method [234] [235], which is widely used for FEP calculations. In umbrella sampling, a potential is added to the Hamiltonian to ensure more uniform sampling along a reaction coordinate. In practice, one needs multiple potentials (windows) to span the gap between the initial and the final states, obtaining overlap of the sampling volume between neighboring windows. The Weighted Histogram Analysis Method (WHAM) [236, 237] is then typically used to combine and unbias the sampling data from the different windows in an optimal manner to reconstruct a PMF. More detail about the umbrella sampling and the WHAM algorithms can be found in Refs. [237], [238] and]239]. Umbrella sampling is perhaps the most widely used approached to compute the free energy of a biomolecular system along a well-defined reaction coordinate and is especially applicable to the study of DNA–protein conformations [110, 111].

Another common approach to free energy calculations is the Adaptive Biasing Force (ABF) sampling method proposed by Darve and Pohorille [240, 241]. In contrast to umbrella sampling which uses a biasing potential, ABF uses a biasing force to drive the simulation along a given path. A detailed and instructive description of the ABF method can be found in the following recent review, Refs. [242] and [243]. Among many applications of these MD sampling methods are the free energy calculations of base flipping and mismatches [85, 86, 89, 90]. The bias sampling techniques are suitable for studying rare events as they focus sampling to the configuration space of interest. Although both algorithms serve a similar purpose, ABF may be more versatile and efficient [242].

The above enhanced sampling algorithms only sample one given path along the free energy surface. However, the transition path from A to B is often not known. Finding an optimal transition path requires sampling free energy along all possible trajectories connecting the initial and final states, for example, to compute a reaction rate constant [244] or to determine a pathway of conformation change [116,117]. One such method is Transition Path Sampling (TPS) [244–246], which is commonly used to study rare events such as DNAbase flipping [91, 109, 115]. Starting from a chosen state along an initial path, one perturbs the momentum of that state and integrates the equations of motion forward and backward in time to generate a new trajectory. The new trajectory is accepted when it connects both initial and final states of the unperturbed system. The initial path may be obtained from a biased MD simulation. Following that, the path can be optimized using path annealing during TPS [91]. More information about TPS can be found in Refs. [247], [248], [249] and [250]. TPS has been instrumental in the study of DNA synthesis and repair [52, 109, 114, 115]. The field of enhanced sampling continues to develop new algorithms, such as selective integrated tempering [251], which has already been applied to study DNA base-flipping processes [88].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.6 Stochastic Model

Among many possible approaches to modeling stochastic dynamics of a biomolecular system, the Markov State model (MSM) is by far the most popular as it enables the study of long timescale dynamics of large biomolecules. A Markov state model is predicated on the assumption that the probability of transitioning between states is determined solely by the current state of the system and not its history. To build a faithful MSM, one first needs to define the ensemble of microstates, which can be done by discretizing the conformational phase space into many sub-states based on certain metrics, such as the root mean square deviation. By counting the number of transitions from state i to state j at a given lag time Δt one can build a transition probability matrix T where each matrix element T_{ij} represents the probability of transitioning from state i to j after a lag time Δt. In this manner, one can investigate the comparatively long timescale dynamics by propagating the transition probability matrix in time according to the following equation:

$$P(N\Delta t) = T^N P(0), \tag{4}$$

where $P(t)$ is the probability distribution of the Markov states at a given time t. By diagonalizing Eq. 4, one can obtain the stationary state associated with the first eigenvalue. The MSM method is described in retails in Ref. [252]. The MSM approach was used to model DNA base pair mismatch repair process [141] and to study the dynamics of the thymine DNA glycosylase [62, 63]. Even though the MSM is able to simulate long-time dynamics, obtaining reliable transition matrix elements requires extensive configurational sampling through MD simulations.

3 Computer Simulations of DNA Replication and Repair Machinery

All cells have enzymes called DNA polymerases that use the base pairing property of DNA to produce a complementary DNA strand from a parent template strand, allowing a cell to replicate and to repair its genome. Base pairing also allows complementary DNA strands to self-assemble into a double helix structure, keeping redundant copies of the genomic information in close proximity and protecting this information from chemical and enzymatic degradation by burying it in the interface of the basepairs. After DNA replication but before cell division, a second copy of each chromosome provides further information redundancy. Protecting the genome is important for all cells and, consequently, several distinct repair pathways have evolved that can be broadly categorized into recombination-dependent and recombination-independent pathways.

3.1 Replication

Origin of replication.—In eukaryotic and prokaryotic cells, DNA replication begins with an initial, protein-mediated unwinding (melting) of a designated part of the genome—the origin of replication, a 200 to 500-basepair stretch of DNA rich in relatively weak adenine-thymine (AT) basepairs. The DNA unwinding is associated with negative supercoiling and with the oligomerization of a helical protein complex (DnaA in bacteria; ORC proteins in eukaryotes) that is believed to mechanically melt the DNA strands [9]. The Zacharias group used all-atom Hamiltonian replica exchange umbrella sampling simulations to explore the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

impact of underwinding a duplex DNA containing a central regions rich in either AT or GC basepairs [112]. The authors found that the AT-rich DNA would melt when the average twist per basepair was reduced to 25–26° at a free energy cost of ~ 1.5kcalmol^{-1} per basepair. Takada *etal.* used a coarse-grained model developed in their lab to simulate the self-assembly of a DnaA helix on DNA in an attempt to solve the structure of the DnaA–DNA oligomer, believed to aid melting of the DNA at the origin, Fig. 2B [155]. The study demonstrated that a flexible linker between oligomerization and DNA-binding domains of DnaA was required for complete DnaA oligomerization. An atomistic model of the DnaA-DNA complex was constructed by mapping into the final configuration of the coarse-grained model, and the stability of the model was demonstrated through brief all-atom simulation.

After the unwinding event, single-stranded DNA binding proteins stabilize the melted DNA and help recruit replication proteins. A protein complex called a replication fork (or replisome) then assembles at each of the two sides of the unwound DNA, at the junction between single-and double-stranded DNA, Fig. 2A,C. Within each replication fork, a hexameric ring-shaped helicase hydrolyses ATP to unwind the double-stranded DNA, driving the complex away from the origin of replication. A protein called primase begins replication on each of the daughter strands by synthesizing a short RNA strand, a primer, that is subsequently extended by a DNA polymerase protein.

Hexameric helicase.—The replicative hexameric helicases in bacteria and eukaryotes contain RecA and AAA+ structural motifs, respectively, that consume ATP at the protomer–protomer in- terfaces to drive cyclic conformational changes that push one of the DNA strands through the central pore of the enzyme leaving the other strand at the enzyme's periphery, which unzips the DNA duplex. The Warshel and Brooks groups independently used coarse-grained modeling to demonstrate how the conformational changes driven by ATP consumption causes translocation—and therefore unwinding—of the DNA [157,158]. Although the methods used by the two groups differed substantially, both studies arrived at the conclusion that electrostatic interactions between the DNA and the helicase facilitate translocation. Furthermore, the Brooks group was able to show that sequential binding of ATP around the hexamer caused translocation in the expected direction and that reversing the sequence of ATP binding reversed the direction of translocation, see Fig. 2D. Neither study, however, attempted to uncover the mechanism governing sequential binding of ATP. More recently, Ma and Schulten used enhanced sampling simulations to drive an atomistic model of two subunits of an RNA helicase related to the bacterial DNA helicase through its operation cycle. They demonstrated that the transition could not proceed when an "old" ADP product was still bound to the second subunit, suggesting that al-losteric interactions mediated by the ADP enables the sequential triggering of ATP consumption to orchestrate the directional transport.

Single-stranded DNA binding proteins.—Single-stranded binding proteins (SSBs; RPA in eu-karyotic cells) are essential components of the DNA replication and repair machinery. The main apparent function of SSBs is to sequester exposed single-stranded DNA that might otherwise form secondary structures (*e.g.* hairpins) or become damaged by free radicals or cut by nucleases. Efficient sequestration of DNA requires strong binding

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

affinity, which is at odds of the other requirement of SSBs to be readily dislocated along the DNA allowing replication and repair proteins to perform their functions. Furthermore, SSBs interact with a number of replication and repair proteins, likely helping them to colocalize with sites of replication and repair and perhaps playing a more subtle regulatory roles.

Coarse-grained simulations were used to predict the structure of SSB-ssDNA complexes in six organisms [156]. In general, the DNA strands were found to be stabilized by a mixture of electrostatic and co-planar stacking interactions. For the case of *Escherichia coli* SSB, our own all-atom simulations support this view [45]. We found that patches of the SSB surface rich in arginine residues interact with the DNA through both electrostatic and stacking interactions, contributing to the stability of the complex. Furthermore, our microseconds-long equilibrium all-atom simulations of the complex, Fig. 2E, revealed spontaneous formation and resolution of small bulges of ssDNA on the SSB surface [45], consistent with experimental findings that suggest that SSB diffuses along DNA (at high rates considering the large affinity of SSB for ssDNA) by the propagation of similar-sized bulges around the SSB surface [254]. Finally, the Anindya group recently generated a model of repair protein AlkB docked onto SSB, demonstrating the stability of the complex using the all-atom MD method [71]. The study showed a route for nucleotides to access the AlkB binding pocket, even when the ssDNA strand forms a complex with SSB.

Processive DNA polymerase.—DNA polymerase enzymes catalyze the triphosphate hydrolysis reaction to extend a nucleic acid strand, a reaction that was recently examined using QM simulations that revealed how Mg^{2+} ions lower the reaction energy barrier [145]. Computational studies of replicative DNA polymerase enzymes are so far limited in number. To our knowledge, the primary replicative polymerases (Pol δ in eukaryotes, Pol III α in bacteria) have not yet been computationally studied, but a pair of studies have investigated the structure of Pol α, which is related to Pol δ and is responsible for the initial extension of the RNA primer [33, 34]. It was found that the “thumb” domain, which cradles the RNA/DNA duplex, adopts a conformation that is induced by the RNA/DNA duplex and is quite distinct from the conformation adopted when not bound to the substrate (the apo state) [34].

Sliding clamps.—One perhaps surprising feature common to many DNA polymerase proteins is that they readily fall off of their template DNA strands. The apparent low affinity of polymerases for template DNA is likely to have been selected to deal with atypical DNA such as abasic sites or sites containing bulky lesions resulting from a chemical damage. When such a site is encountered, the main replication polymerase falls off the DNA while remaining tethered to a ring-shaped “clamp” protein that circumscribes the DNA [255]. A slower but more error tolerant polymerase also interacting with the clamp can then take over replication at the lesion [256]. The clamps are also known to interact with a wide variety of DNA replication proteins, coordinating the many subtle aspects of DNA replication. For example, at the end of Okazaki fragments, clamps recruit exonuclease to digest the RNA primer (in eukaryotes) and DNA ligase to fix nicks in the newly synthesized DNA strand.

The replication fork contains a spiral-shaped protein, the clamp loader, that deposits clamps near every RNA primer. All-atom and coarse-grained simulations were used to study the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fluctuations of both open and closed clamps [39,152]. Additionally, external forces have been applied to induce rupture of a clamp [40]. These studies demonstrate how sliding clamps are able to open at oligomer interfaces and that fluctuations of the open clamp are large enough to match the helical pitch of the clamp loader. In bacteria, the clamp loader also directly contacts the helicase and DNA polymerase molecules, acting as a central hub for replication.

Removal of RNA primers.—Every 100–1000 nt, the DNA polymerase on the lagging strand encounters an RNA primer. In bacteria, Pol III is typically exchanged for Pol I, which synthesizes DNA to replace the primer while degrading the primer one base at a time. In eukaryotic cells, the polymerase displaces the primer, after which the FEN1 endonuclease (RNase H in some species) cleaves the primer. The conformational fluctuations of FEN1 bound to the PCNA clamp were observed in all-atom simulations that demonstrated that the endonuclease adopted several conformations in the absence of DNA, but just a single conformation when bound to DNA [67] The catalytic activity of RNase H and FEN1 has been examined through several QM/MM studies [131–133]. The consensus from these studies is that the two endonucleases utilize similar mechanisms for catalyzing phosphodiester cleavage, and that the rate limiting step in catalysis is the deprotonation of a single water molecule coordinated by the enzyme.

After primer removal, DNA ligase I repairs the resulting nick. Although we are not aware of computational studies of the ligase itself, a computational study from the Tomkinson and MacKerell groups searched for potential inhibitors of a ligase using a combination of docking and MD simulations [128]. Because cancer cells proliferate more rapidly than ordinary cells, inhibiting enzymes involved in replication and repair processes is a possible target for new therapeutics.

Mitochondrial replication.—The ~ 17,000 basepairs of the circular DNA in human mitochrondria is replicated by DNA polymerase γ, one strand at a time [257]. The DNA polymerase γ was studied by the Sharma and Suomalalnen groups through microseconds-long all-atom simulations. The simulations revealed a number of structural transitions that occur upon DNA binding, including exposure of the catalytic sites [38]. These results are significant in part because over 200 known mutations of this polymerase can cause metabolic diseases.

3.2 Repair

DNA is susceptible to damage originating from a wide range of sources including mistakes during DNA replication, oxidation by free radicals, and reactions stimulated by ionizing radiation, including UV light. The typical products of DNA damage include basepair mismatches, abasic or chemically modified nucleosides, single-stranded breaks, intra-and inter-strand crosslinks, and double-stranded breaks. The detection of severe DNA damage can trigger apoptosis, a programmed cell death, which helps prevent unwanted mutations. Many cancers are associated with defects in DNA repair mechanisms and/or damage detection signaling pathways, allowing the accumulation of mutations that result in unchecked growth and replication [258]. Hence, cells have evolved several distinct repair

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mechanisms to cope with different kinds of DNA damage [259], which are schematically summarized in Fig. 3. In general, cells prefer repair processes that use undamaged DNA as a template during the repair to prevent the introduction of mutations. The availability of such template DNA, and therefore the chosen DNA repair pathway, depends on the type of damage and on the phase of the cell cycle. The most important kinds of DNA repair include: base excision repair, nucleotide excision repair, mismatch repair, template-directed repair, translesion synthesis, non-homologous end-joining (NHEJ), and repair through homologous recombination (HR).

This section is organized as follows. First, we describe what simulations have revealed about the main error-correcting processes that occur during DNA replication (Section 3.2.1). Next, we discuss how damage to single-stranded DNA is detected and how repair is initiated by the three main single-strand damage repair pathways (Section 3.2.2). We then describe how polymerases complete the repair of damage to single-stranded DNA (Section 3.2.3). Finally, we review what a few simulations have shown about the repair of double-stranded DNA breaks (Section 3.2.4).

3.2.1 Repair of errors encountered during DNA replication—There are two kinds of errors that should be repaired during DNA replication. First, the DNA polymerase can fail to recognize the identity of a base on the template strand, which would result in a mismatch if not for the proofreading exonucleolytic domains on the key replicative polymerases. Second, lesions might be encountered on the DNA that were unrepaired during the Gap phase of the cell cycle preceding synthesis that the replication fork must be able to move past using translesion synthesis in order to prevent serious chromosome errors.

Proofreading during DNA synthesis.: DNA polymerase proteins involved in replication have a small chance of incorporating a mismatched basepair on the first pass. In both eukaryotic and bacterial cells, most polymerases (including the main replicative polymerases) possess ex-onucleolytic domains that perform proofreading on the nascently synthesized DNA as a first line of defense against replicative errors. We are unaware of any computational studies of the catalytic activity of these proofreading domains, but several all-atom MD studies have demonstrated what appears to be a lack of direct coupling between the polymerization domain and the exonuclease domain [38, 42, 43]. The simulations suggest that the enzymes switch between polymerizing and proofreading states on roughly the microsecond timescale rather than performing both functions simultaneously.

Translesion synthesis.: DNA damage encountered during replication can be risky to repair because the duplex is unwound and chromosome defects can readily result from a failed attempt to repair the DNA. When the processive DNA polymerase stalls due to DNA damage, an error-tolerant DNA polymerase can attempt to move past the error and restart the replication process. This process, called “translesion synthesis”, is carried out by Pol η, κ, ι, and ζ in eukaryotic cells, and Pol IV and V in bacteria. These polymerases replicate undamaged DNA with low fidelity, but each can recognize specific types of DNA damage and synthesize a correct basepair despite the damage with high fidelity compared to the processive polymerases. A common bulky lesion that stalls the replicative polymerase is an intrastrand thymine–thymine dimer resulting from UV radiation. All-atom MD simulations

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

showed that Pol η adopts the same overall conformation regardless of the presence of a thymine–thymine dimer on the template strand and regardless of whether the incoming nucleotide is a dATP or mismatched dGTP molecule [56]. Rather, it was shown that the enzyme's binding region has more hydrogen-bonds with the thymine dimer than with undamaged thymines, increasing the rigidity of the damaged motif. The enzyme appeared to have relatively low sensitivity to the incoming nucleotide, consistent with experimental observation. Some of these features may be specific to the type of DNA damage or to the polymerase; MD simulations of Pol IV in complex with a DNA molecule containing a guanine with a bulky benzene adduct showed that the polymerase stabilized the nucleotide but had only limited interactions with the adduct, allowing it to move out of the base-stacking plane and into the major groove [44]. This motion exposed much of the Watson-Crick surface of the damaged guanine for basepairing with an incoming dCTP.

QM/MM simulations of Pol IV indicated that the reaction pathway involves the transfer of a proton from the 3' hydroxyl group of the nascent strand to water molecules in solution, allowing the 3' oxygen to bind the first phosphate in the incoming dCTP molecule [134]. The reaction was associated with a 20 kcal mol^{-1} barrier that was higher than the reaction barrier observed for the more accurate Pol β. The higher barrier was attributed to greater exposure to solvent in the more open active site of Pol IV, which can more easily accommodate damaged template DNA compared with Pol β.

Cell cycle checkpointing.: DNA replication cannot be stopped once it has started. If the DNA polymerases are unable to replicate past a lesion, the helicase may continue to unwind DNA resulting in the exposure of significant amounts of ssDNA. In eukaryotic cells, the exposed DNA is coated by the single-stranded DNA binding protein RPA, which interacts with the protein kinase ATR that causes a cascade of phosphorylation signals including checkpoint proteins Chk1 and Chk2, increasing expression of proteins involved in repair, homologous recombination (HR), and apoptosis—programmed cell death. Activation of the checkpoint proteins also stalls the progress of the cell cycle, giving the cell more time to repair its DNA and resume replication [260].

3.2.2 Detection of single-stranded DNA damage—

Damage to single-stranded DNA takes many forms and is repaired primarily through three mechanisms, as summarized schematically in Fig. 3. Base excision repair is used to remove a wide variety of spontaneously occurring DNA damage to DNA bases and is also used to demethy-late a base, removing a common epigenetic marker. Bulky DNA lesions, for example resulting from exposure to UV radiation, are normally repaired by the nucleotide excision repair pathway. Errors during DNA replication and also spontaneous chemical reactions can result in mismatched basepairs, which are repaired by the mismatch repair pathway.

All three of these mechanisms use proteins that mechanically detect damage on the DNA. In mismatch repair, the MutS homologue (MSH; MutS in bacteria) bends DNA to detect mismatches. In base excision repair, a number of unique glycosylase proteins each detect a single kind of DNA damage. Finally, in nucleotide excision repair, theXPC/Rad23B complex (UvrA/UvrB in bacteria) detects distortion of the DNA indicative of damage. All

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

damage detection mechanisms of all three repair pathways rely in part on the altered stability and elastic properties of damaged DNA.

Energetics of damaged and mismatched basepairs.: *Ab initio* simulations have been used to compare the energetics of intact and mispaired Watson-Crick basepairs [142–144, 147, 148], demonstrating that selection of natural basepairs is primarily driven by hydrogen bonding and solvent interactions rather than stacking. A large number of MD studies have characterized the energetics involved in bending and flipping intact, damaged and mismatched bases of various sequences out of a DNA duplex [83–89, 92–94, 105, 107], a topic that was previously reviewed [105]. The simulations generally show that bases flip through the major groove, that the free energy cost of flipping is lower for mismatched bases compared with intact ones, and that the transition state is generally unfavorable, causing a slow kinetics for switching between intrahelical and extrahelical states. Unfortunately, the studies have not comprehensively covered the space of likely forms of DNA damage, and it remains to be determined whether general rules for flipping damaged bases can be established.

The Feig group used all-atom enhanced sampling simulations to quantify the free energy required to bend intact and mismatched nucleotides in electrolyte solution [90], Fig. 4 A. The free energy cost of bending duplexes beyond ~ 60° decreases with the introduction of a mismatch, regardless of the nucleotide sequence. Duplexes typically bent towards the major groove, widening the minor groove. Duplex bending occurred smoothly for canonical B-DNA, but with a mismatch, the duplex became abruptly kinked. Finally, the stability of mismatches was found to rank as follows, from most to least stable: pyrimidine-purine, purine–purine, pyrimidine–pyrimidine. These structural and energetic features likely facilitate initial recognition of mismatches by the repair protein (MutS) as it scans along the genome.

Damage detection in mismatch repair.: The detection of mismatched basepairs is performed by the MutS protein. MutS has a ring-like dimeric structure where each mismatch-binding domain is separated from another by the ATPase domain, a distance of about 7 nm. All-atom MD simulations have been used in attempts to understand how MutS recognizes mismatches and how changes in the mismatch binding domain are coupled to changes in the ATPase domains [76–82]. The Feig group revealed a network of interactions that suggested the following overall mechanisms of MutS activity [79], Fig. 4 B. First, non-specific DNA binding induces dissociation of one of the two ADP molecules likely to be bound to MutS in the cellular environ-ment [261], allowing the protein to slide along the DNA. Next, encountering a mismatch results in an exchange of the remaining ADP molecule for ATP, followed by ATP binding to the second ATPase domain and rapidly hydrolyzing to form ADP. It is presumed that this ATP–ADP-bound configuration initiates the repair process.

Damage detection in base excision repair.: Base excision repair begins with the detection of DNA damage by a DNA glycosylase, which wraps around half a duplex, distorting one strand to expose its base. There are many DNA glycosylase proteins, each able to detect one kind of a DNA base damage by flipping the base into a highly-selective base-binding pocket.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

If a damaged base is found, the glycosylase cleaves the bond connecting the base to the deoxyribose, initiating the repair process.

The glycosylase proteins uracil DNA glycosylase [61], thymine DNA glycosylase [62, 63] and MutM [141] have been studied through atomistic MD simulations. In particular, the base flipping process in thymine DNA glycosylase was studied by first driving the DNA-protein complex from an intrahelical to extrahelical state, which was followed by 300 100-ns simulations all initialized along the base-flipping pathway for a total of 30 μs [63], Fig. 5. The results of these simulations were used to construct Markov state models describing the kinetics of transition between 500–800 microstates that characterize the base-flipping process. The model revealed a chain of six metastable states that required tens to hundreds of μs to transition between. The protein was found to stabilize early opening of the damaged basepair, facilitating the transition and inserting an arginine sidechain in the space left by the vacated base. Similar observations were made regarding the path and energetics of base flipping in MutM using enhanced sampling simulations [75].

Surprisingly, the simulations also uncovered a preference for the flipping transition to occur through the minor groove, in stark contrast to the results of the solution simulations of duplex DNA [90]. Why would nature select the more difficult pathway? We tentatively speculate that DNA damage could decrease the free energy barrier for flipping through the minor groove more than for flipping through the major groove. Additional studies would be required to verify this hypothesis.

QM/MM simulations were used to demonstrate that the energy barriers for base cleaving were lower when a ribose-protonation pathway was assumed compared to a base-protonation pathway favored in older literature [61]. After protonation of the ribose, glycosydic cleavage can occur readily because there is almost no free energy barrier. One implication of that study is that the cleavage reaction is independent from the type of the base and could occur even for intact bases. Hence, the study suggests that base discrimination occurs during the flipping step, not the reaction step.

After cleavage, the base must be released from the glycosylase's binding pocket. The same researchers that built a Markov state model of base flipping in thymine DNA glcosylase used similar methods to study the release of an excised base from thymine DNA glycosylase. The simulations revealed a two- or three-step release process whereby the base is partially displaced from the binding pocket by water molecules before it is released into a solvent-filled channel. The overall release process is expected to occur in just 10 μs.

Damage detection in nucleotide excision repair.: Bulky DNA lesions resulting, for example, from exposure to UV radiation, are normally repaired by the nucleotide excision repair pathway [262]. First, the protein complex XPC/Rad23B (Rad4/Rad23 in yeast; UvrA/UvrB in bacteria) probes for DNA damage. Unlike glycosylase proteins, of which there are many variants that detect different kinds of damage, this complex is responsible for detection of all kinds of DNA damage repaired through the nucleotide excision pathway. All-atom umbrella sampling simulations were used to explore a number of possible DNA-damage recognition pathways by Rad4 [72, 73, 263]. Based on the simulations, the authors

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

suggested the following mechanism of damage recognition, Fig. 6: (i) a DNA binding domain tethers a pair of DNA damage recognition domains to a helix, allowing them to probe for unusual DNA flexibility; (ii) when a lesion is encountered, a closer association of the damage detection domains correlates with unwinding and bending of the DNA duplex; (iii) nucleotides paired with the lesion are flipped into a pocket of one of the DNA binding domains; and (iv) the damage recognition domains insert a β-hairpin, exposing the lesion. The simulations show that the third step requires overcoming a 4–5 kcal mol^{-1} free energy barrier. The simulations suggest that the protein is able to recognize a broad range of DNA damage by avoiding direct contact with the damaged bases, instead taking advantage of the reduced stability of damaged basepairs to force bases on the damaged strand into solution where they can be recognized and repaired by other proteins.

3.2.3 Nucleases, polymerases and ligases in the excision repair processes—

After detection of damage to single-stranded DNA and associated excision of the damage, DNA polymerases must synthesize new DNA to replace the damaged bases using the undamaged complementary strand as a template. Below, we review the mechanisms that are used to cut, synthesize, and join DNA to complete the excision repair processes. Here, we omit a description of the completion of mismatch repair, which is usually performed by the replicative polymerases Pol δ in eukaryotes and Pol III in bacteria.

Base excision repair.: Following base excision, an apurinic endonuclease cuts the DNA backbone (some glycosylase proteins perform this function). The interaction of human apurinic endonuclease with DNA was studied using all-atom MD simulations [68]. The simulations revealed significant mechanical deformation of the damaged DNA duplex by the endonuclease, a mechanism which might prevent of cleavage of undamaged DNA. After the endonuclease cuts the DNA backbone, a DNA polymerase, typically Pol β in eukaryotes, performs “short-” or “long-patch” base excision repair, where a single nucleotide or 2 to 10 nucleotides are synthesized to replace the missing base. In long-patch base excision repair, the action of the polymerase displaces a stretch of the original DNA strand, and the exposed ssDNA is cleaved by a flap endonuclease. Finally, in either repair pathway, a DNA ligase repairs the break in the DNA strand left by the polymerase. In the short-patch pathway, repair is partially coordinated by the ssDNA break repair protein, XCCR1, which binds Pol β, DNA ligase III, and the ssDNA break binding protein PARP1.

Pol β is the best computationally-studied polymerase [120]. Most studies have explored the energetics and possible pathways of the reaction for intact basepairs [46, 48, 115, 117, 119, 120, 123, 125, 129, 136, 138]. A few studies have explored the relationship between large-scale conformation of the repair polymerase and its fidelity [47, 55]. The Schlick group has shown that the reaction pathway strongly depends on the initial geometry of the coordinating ions [116].

Nucleotide excision repair.: Pol I, which synthesizes DNA in nucleotide excision repair in bacteria, has been the subject of several all-atom simulation studies [28, 29, 32]. Forces were applied to convert the polymerase between open and closed states, a motion that may be the rate limiting step of polymerase kinetics. The closed conformation aligns the incoming nucleotide so that it pairs with the template strand. Hence, the transition to the closed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conformation can be expected to be important for preventing mismatches. All-atom simulations were used to estimate the dependence of the kinetics of the transition on the tension in the template strand to predict the rate of replication [32]. Microsecond-long unbiased simulations were used to examine the transition between open and closed conformations, revealing a network of tertiary interactions within the protein that facilitate the transition [28] and revealing how a magnesium ion coordinates the incoming nucleotide [29]. Additional simulations revealed that the presence of an incorrect incoming nucleotide biases the polymerase towards the inactive open configuration [108].

3.2.4 Repair of dsDNA breaks—If a dsDNA break is detected during the DNA synthesis or early mitosis states of the cell cycle, it is usually repaired by homologous recombination (HR) with the sister chromatid used as a template. During the rest of the cell cycle, a homologous stretch of dsDNA is unlikely to be immediately available and, thus, the dsDNA breaks are typically repaired by a more error-prone non-homologous end joining (NHEJ) pathway, as described below. Because of its important role not only in genome maintenance, but also in promoting genetic diversity of a species, we delay discussion of HR until the next section.

Non-homologous end joining.: Most bacteria lack repair through NHEJ, but in the past fifteen years, a few bacterial genera have been shown to possess minimalist NHEJ mechanisms that are so far known to rely only on two proteins, Ku and LigD, a DNA ligase that has a polymerase domain. From studies of eukaryotic cells, the Ku70/80 heterodimer is known to detect double-stranded DNA breakage and encircle each end of the broken double helix, likely forming protein-protein contacts across the break to hold the strands together while other proteins catalyze repair. We used all-atom MD simulations to show that the standard binding free energy for stacking blunt-ended DNA molecules is ~5 kcal mol^{-1}, an energy large enough to align broken DNA ends provided proteins (such as Ku) keep the DNA ends in close proximity [60]. All-atom MD simulations probed binding and unbinding of DNA helices of different nucleotide sequences to Ku70, estimating the binding energy [59]. Although the simulations were very brief, it was observed that the Ku70 ring contracts without Ku80, and that interactions with the ring dominate the enthalpy of DNA binding, likely facilitating directional loading of Ku. Using docking simulations, the same researchers proposed a pair of binding geometries for Ku80 bound to Ku70 or DNA, complexes that have not yet been crystallized [126].

In eukaryotic cells, Ku interacts with a number of other proteins that are involved in cell-cycle checkpointing and that facilitate the repair of the break by DNA ligase IV and Pol λ. Of the other proteins involved in NHEJ, we are only aware of computational studies that searched for drugs that inhibit DNA ligase IV for possible oncological applications [264], and of a study of Pol λ that showed how a single polymorphism impacts the geometry of DNA bound to the protein [57]. Cell-scale kinetic modeling of the NHEJ repair process has been performed by several groups [170, 171, 186, 187]. In general, these simulations seek to relate the kinetic rates of individual proteins involved in NHEJ with overall DNA repair rates after exposure to DNA damage.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Decision to use NHEJ versus HR.: The decision to use NHEJ or HR depends on the presence of a DNA-binding complex with exonucleolytic capability, Mre11-Rad50-Xrs2 (MRX) in yeast or Mre11-Rad50-NBS1 (MRN) in mammals. Ku is expressed throughout the cell cycle, but MRX/MRN are expressed mainly between S and G_2, when a sister chromatid is likely to be available [265]. During G_1, the lack of an MRX/MRN complex ensures that NHEJ is the chosen repair pathway for double-stranded breaks.

3.3 Recombination

Efficient DNA repair mechanisms are alone insufficient for the continued existence of a species because of environmental changes over time. As a result, all life forms have also developed a mechanism for evolution, which allows them to adapt their genetic information to changes in the environment. Amazingly, the homologous recombination (HR) process used to repair double-stranded DNA breaks is also the primary mechanism for promoting genetic diversity [266, 267]. A recombinase protein, in concert with other enzymes, works to swap DNA between two DNA molecules of similar DNA sequences, Fig. 7. In the context of a double-stranded DNA break, the process can repair a chromatid without introducing dangerous mutations. In the context of reproduction, the process can result in mixing of the genes present on a parent cell's two chromatids.

The process of homologous recombination involves the following three major steps: the search for the right fragments of the partner DNA molecule (homology recognition), exchange of the DNA strands between the pair of DNA molecules (strand exchange), and resolution of the branched DNA structure (branch resolution). Both homology recognition and strand exchange are mediated by the RecA protein in bacteria (Rad51 in eukaryotes). Fig. 8 A shows a molecular graphics image of a RecA protein. In the presence of fuel molecules of adenosine triphosphate (ATP), RecA polymerizes on single-stranded DNA-the reference DNA strand—to form a right-handed filament, Fig. 8 B. This RecA filament has a remarkable ability to search for the homologous fragment of DNA in the crowded environment of a biological cell and facilitate the transfer of one DNA strand from the partner DNA molecule to the reference DNA strand without additional energy input (i.e., ATP hydrolysis).

Homology search.—Several crystal structures of a RecA filament bound to single-and doublestranded DNA have been published [268]. The crystal structures have shown each RecA monomer to be in contact with three contiguous nucleotides—a nucleotide triplet—of the reference DNA strand, Fig. 8 A. Each triplet has a conformation similar to the conformation of a nucleotide triplet in a DNA double helix, but the loops of the RecA protein separate the neighboring triplets into distinct groups of nucleotides, Fig. 8 A,C. Due to the gaps between the groups of triplets, the reference DNA strand in a RecA filament is extended by about 50% in comparison to that in a canonical double helix. It is generally presumed that stretching of the ssDNA facilitates the search for homology, allowing the filament to quickly scan dsDNA for a homologous se-quence. Coarse-grained MD simulations determined that stretching every third pair of phosphate groups in a DNA molecules to the same extent as in the RecA-DNA complex causes hybridization and dehybridization to occur cooperatively in three-nucleotide steps [161]. The DNA triplets

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were seen to have reduced stability compared with those of B-form DNA, which might accelerate the search process.

Several numerical models have been used to study the homology search process [172, 173, 188]. In general, the models assume some non-specific attraction between the Rad51–ssDNA filament and the dsDNA being searched. One of the models reviewed here considers a step size for the search that matches the size of the filament [173]. In that model, the rates for the random walk performed by the filament depend on the filament size. The model predicts that the shorter the Rad51–ssDNA filament, the longer the homology search will take. By contrast, the models from Fulconis, Dutreix and Viovy discretize the search into sites as small as 20 basepairs [172, 188]. Their latest model predicts rapid extension of the RecA after a slow nucleation step. Interestingly, that model takes into account the DNA sequence and shows the possibility for a homology search to become kinetically trapped by contacting a partially homologous DNA segment.

Another intriguing possibility is that the force between two double-stranded DNA molecules can depend on homology of their nucleotide sequence [269, 270]. Recent all-atom MD simulations and single-molecule experiments [99, 100] provide some support to this idea by showing that DNA molecules rich in AT basepairs exhibit stronger attraction than CG-rich DNA duplexes in the presence of multivalent cations such as spermine and polylysine. Interestingly, the interaction strength was found to be modulated predominantly by the methyl groups of the T bases, predicting stronger attraction between methylated DNA fragments and partially explaining why double-stranded RNA (where T is replaced by U) resists condensation [271]. Since the nucleus is likely to contain many polyamines in the form of lysines on histone tails, it is plausible that the greater attraction for AT would promote alignment of homologous sequences, making the homologous search more efficient.

Structural modeling of the Rad51 and RecA filaments.—During the strand exchange process, the Rad51–ssDNA filament binds dsDNA, partially melts or otherwise exposes the bases of the dsDNA. Subsequently the exposed dsDNA bases pair with the stretched ssDNA in the filament, facilitating strand exchange. Crystallographic structures that would elucidate the detailed sequence of events during this process have not been available. Hence, several simulation studies have sought insights into the structure of the RecA and Rad51 filaments.

Prior to the availability of a crystallographic structure of a Rad51 helix in an active state, homology modeling and all-atom simulations were used in an attempt to solve the dimer structure of the Rad51 filament in its active state [97]. The simulations showed the importance of counterions near the ATP binding site at the junction of the monomers for maintaining interactions at the dimer interface, and predicted a model for the filament.

Regulation of RecA and Rad51 filamentation is important for cells to achieve a balance between replication and repair/recombination activities. One regulatory mechanism is the inhibition of RecA filamentation by direct interaction with the RecX protein [272]. All-atom MD simulations were used to complement small-angle neutron scattering data to elucidate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the structure of RecX interacting with a RecA filament [95]. The simulations indicated that RecX has very little direct interaction with RecA, but has extensive electrostatic interactions with the ssDNA. These interactions may limit the ability for RecA to polymerize on ssDNA.

It has been suggested that the disordered acidic tails of RecA might interact with the DNA binding sites to exert some regulatory effect. Intrinsically disordered regions of a protein are rarely solved by crystallography and their myriad conformations are notoriously difficult to sample in atomistic simulations. Nevertheless, a simulation study complementing experiments included the tails in an effort to determine how they interact with the filament [96]. Although the conformations of the tails could not be extensively sampled, the all-atom simulations indicated that the tails do not directly interact with the dsDNA binding site—a suggested regulatory mechanism. However, interactions were seen between the tails and the RecA core at a likely "gateway" for dsDNA entering the filament. Taking simulation and experimental results together, the authors suggested that increased Mg^{2+} concentration could disrupt the interactions between the tails and the gateway, opening the path for dsDNA binding.

Coarse-grained docking simulations were used to search for an encounter complex between the RecA–ssDNA filament and dsDNA [166]. Two dsDNA structures were considered, one in canonical B-form, the other highly curved. The docking simulations indicate that dsDNA encounters the filament at an angle that locally approximates the helical pitch of the filament. The geometry of the docked dsDNA molecule included some minimal contacts between the bases of ssDNA and the major groove of the dsDNA. Hence, the results of the docking simulation support a model for homology search in which the ssDNA bases open the dsDNA molecule from the major groove to test for homology rather than one where the RecA protein first denatures the dsDNA molecule.

Branch migration and resolution.—Once the search for homologous DNA is completed and strand invasion has occurred, a DNA polymerase (usually Pol δ or Pol I) synthesizes DNA to extend the 3' ends of the broken DNA. The resulting four-strand DNA intersections, known as Holliday junctions, are then able to passively diffuse, stochastically affecting the final locations where DNA is exchanged between the chromatids. The structure and dynamics of Holliday junctions have been studied through coarse-grained [103, 159] and atomistic [101, 102, 104, 160] simulations. Simulations show that Holliday junctions favor two distinct stacked geometries of differing chirality and that magnesium ions may help stabilize these geometries. An open conformation provides a transition state between these two isomers that may also facilitate branch migration. A recent study from our group [104], however, highlighted the challenges in quantitatively reproducing the experimental observations using the all-atom MD approach, which were attributed to the imperfection of the molecular force field models. In bacteria, branch migration is driven in an ATP-dependent manner by the RuvAB complex. All-atom umbrella sampling simulations suggested that RuvA forces the junction into an open conformation [118].

The main pathway used to resolve the branch intermediate during the repair of double-stranded DNA breaks involves recognition and subsequent nicking of Holliday junctions by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the endonuclease Resolvase A, or RuvC in bacteria. Unfortunately the mechanistic details of the operation of these enzymes remain obscure so far.

4 Outlook

We have described a spectacular array of biological processes responsible for faithful duplication and maintenance of the genomic information. Looking back, the complexity and multiplicity of the processes appear to present a daunting challenge to modeling. From a technical standpoint, the principle difficulties lay in the multitude of time and length scales involved. Being the very essence of the processes, chemical transformations occur at the level of individual DNA nucleotides but involve only a very small fraction of all DNA material present in the system. The replisome complex is a diffuse agglomeration of protein machines whose action occurs along many, sometimes interdependent, reaction coordinates. The replication time scale—milliseconds to minutes—is intractable using brute force approaches. It is, thus, not surprising that the vast majority of computational studies in the field of DNA replication and repair investigated the function of individual biomolecules, corresponding to isolated steps of multi-step, multi-player biological processes.

And yet, we believe the field is ripe for the computational approaches to connect disjoint processes into coherent physical models that are capable of reproducing the DNA replication and repair processes at a quantitative level and predicting the experimental outcomes. Although our knowledge of DNA replication in eukaryotic cells is still incomplete, the replication processes are well characterized for bacterial species and for viruses that infect bacteria. The latter systems can also be reconstituted *in vitro* [273] and thus present ideal initial modeling targets.

The first necessary step toward a realistic computational model of DNA replication and repair phenomena is development of physics-based models that can quantitatively describe the behavior of protein/DNA systems as it emerges from the intricate interactions of the system's components. Specific to DNA replication, such models would need to satisfy the conflicting requirements of being able to handle very large systems at physiological times scale while retaining atomic-level detail in the description of interactions between DNA and proteins. We believe that such models can be built through integration of the simulation methods that describe the same phenomena at multiple length and time scales. Doing so will require asynchronous, ensemble treatment of individual protein–DNA or protein–protein interactions and stochastic treatment of the network of such interactions. Modeling truly *in vivo* conditions would require development of a "thermostat of proteins", where the abundance of molecular species can be regulated according to gene expression levels expected at particular states of the cell cycles. Development of such computational approaches will be non-trivial but potentially very rewarding as they can be adapted to describe other spatially sparse processes in biological cells. We anticipate the DNA metabolism field to become a hotspot for the development of multi-scale modeling in computational biophysics, producing physical models, numerical algorithms, and software that would ultimately lead to the first predictive, physics-based model of an entire biological cell.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgment

This work was supported by the Center for the Physics of Living Cells through a grant from the National Science Foundation [PHY-1430124] and by the Center for Macromolecular Modeling and Bioinformatics through a grant from the National Institutes of Health [P41-GM104601].

References

[1]. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of The Cell. Garland Science, New York & London, 5th edition, 2007.

[2]. Calin GA, Croce CM. Oncogene 2006 25, 6202–6210. [PubMed: 17028600]

[3]. Feuk L, Marshall CR, Wintle RF, Scherer SW. Hum. Mol. Gen. 2006 15, R57–R66. [PubMed: 16651370]

[4]. Finkel T, Holbrook NJ. Nature 2000 408, 239–247. [PubMed: 11089981]

[5]. Lidzbarsky G, Gutman D, Shekhidem HA, Sharvit L, Atzmon G. Front. Med. 2018 5, 104.

[6]. Lindström S, Andersson-Svahn H. Lab Chip 2010 10, 3363–3372. [PubMed: 20967379]

[7]. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. Science 2017 357, eaag0025.

[8]. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Annu. Rev. Biochem 2010 79, 89–130. [PubMed: 20373915]

[9]. Bell SP, Dutta A. Annu. Rev. Biochem 2002 71, 333–374. [PubMed: 12045100]

[10]. Moffitt JR, Chemla YR, Smith SB, Bustamante C. Annu. Rev. Biochem 2008 77, 205–228. [PubMed: 18307407]

[11]. De Vlaminck I, Dekker C. Annu. Rev. Biophys. 2012 41, 453–472. [PubMed: 22443989]

[12]. Roy R, Hohng S, Ha T. Nat. Methods 2008 5, 507–516. [PubMed: 18511918]

[13]. Hohng S, Zhou R, Nahas MK, Yu J, Schulten K, Lilley DMJ, Ha T. Science 2007 318, 279–83. [PubMed: 17932299]

[14]. Comstock MJ, Whitley KD, Jia H, Sokoloski J, Lohman TM, Ha T, Chemla YR. Science 2015 348,352–354. [PubMed: 25883359]

[15]. Huang B, Bates M, Zhuang X. Annu. Rev. Biochem 2009 78, 993–1016. [PubMed: 19489737]

[16]. Blehm BH, Selvin PR. Chem. Rev. 2014 114, 3335–3352. [PubMed: 24666199]

[17]. Dekker J, Marti-Renom MA, Mirny LA. Nat. Rev. Genet. 2013 14, 390–403. [PubMed: 23657480]

[18]. Schlick T, Hayes J, Grigoryev S. J. Biol. Chem. 2012 287, 5183–5191. [PubMed: 22157002]

[19]. Takada S, Kanada R, Tan C, Terakawa T, Li W, Kenzaki H. Acc. Chem. Res. 2015 48, 3026–3035. [PubMed: 26575522]

[20]. Sarai A, Kono H. Annu. Rev Biophys. Biomol. Struct. 2005 34, 379–398. [PubMed: 15869395]

[21]. Staby L, O'Shea C, Willemoës M, Theisen F, Kragelund BB, Skriver K. Biochem.J. 2017 474, 2509–2532. [PubMed: 28701416]

[22]. Ucuncuoglu N, Andricioaei I, Sari L. J. Mol. Graph. Model. 2013 44, 286–296. [PubMed: 23933704]

[23]. Kenzaki H, Takada S. PLoS Comput. Biol. 2015 11, e1004443. [PubMed: 26262925]

[24]. Saito M, Terakawa T, Takada S. Mol. Sim. 2017 43, 1315–1321.

[25]. Jacquemin D, Zúñiga J, Requena A, Céron-Carrasco JP. Acc. Chem. Res. 2014 47, 2467–2474. [PubMed: 24849375]

[26]. Zhang M, Wang L, Zhong D. Arch. Biochem. Biophys. 2017 632, 158–174. [PubMed: 28802828]

[27]. von Sonntag C. In Adv. Quantum Chem, volume 52, pages 5–20. Academic Press, 2007.

[28]. Miller BR III, Parish CA, Wu EY. PLoS Comput. Biol. 2014 10, 1–15.

[29]. Miller BR, Beese LS, Parish CA, Wu EY. Structure 2015 23, 1609–1620. [PubMed: 26211612]

[30]. Golosov AA, Warren JJ, Beese LS, Karplus M. Structure 2010 18, 83–93. [PubMed: 20152155]

[31]. Graham SE, Syeda F, Cisneros GA. Biochemistry 2012 51, 2569–2578. [PubMed: 22397306]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[32]. Andresen K, Das R, Park HY, Smith H, Kwok LW, Lamb JS, Kirkland EJ, Herschlag D, Finkelstein KD, Pollack L. Phys. Rev. Lett. 2004 93, 248103. [PubMed: 15697865]

[33]. Richartz A, Höltje M, Brandt B, Schäfer-Korting M, Höltje H-D. J. Enzyme Inhib. Med. Chem. 2008 23, 94–100. [PubMed: 18341260]

[34]. Perera RL, Torella R, Klinge S, Kilkenny ML, Maman JD, Pellegrini L. Elife 2013 2, e00482. [PubMed: 23599895]

[35]. Wang H, Hong W, Paterson IC, Pu J, Laughton CA. Mol. Sim. 2014 40, 1290–1299.

[36]. Cox K, Watson T, Soultanas P, Hirst JD. Proteins: Struct., Funct., Bioinf 2003 52, 254–262.

[37]. Wang H, Cui J, Hong W, Paterson IC, Laughton CA. J. Mol. Model. 2013 19, 4997–5006. [PubMed: 24068309]

[38]. Euro L, Haapanen O, Rog T, Vattulainen I, Suomalainen A, Sharma V. Biochemistry 2017 56, 1227–1238. [PubMed: 28206745]

[39]. Oakley AJ. PLoS One 2016 11, 1–18.

[40]. Rydzewski J, Strzalka W, Nowak W. Chem. Phys. Lett. 2015 634, 236–242.

[41]. Xu X, Guardiani C, Yan C, Ivanov I. Nucleic Acids Res. 2013 41, 10020–10031. [PubMed: 24038358]

[42]. Xu X, Yan C, Kossmann BR, Ivanov I. Phys J. Chem. B 2016 120, 8379–8388.

[43]. Gahlon HL, Walker AR, Cisneros GA, Lamers MH, Rueda DS. Phys. Chem. Chem. Phys. 2018 20, 26892–26902. [PubMed: 30345999]

[44]. Wilson KA, Wetmore SD. Chem. Res. Toxicol. 2017 30, 2013–2022. [PubMed: 28810119]

[45]. Maffeo C, Aksimentiev A. Nucleic Acids Res. 2017 45, 12125–12139. [PubMed: 29059392]

[46]. Oelschlaeger P, Klahn M, Beard WA, Wilson SH, Warshel A. J. Mol. Biol. 2007 366, 687–701. [PubMed: 17174326]

[47]. Yang L, Beard WA, Wilson SH, Broyde S, Schlick T. Biophys.J. 2004 86, 3392–3408. [PubMed: 15189842]

[48]. Yang L, Arora K, Beard WA, Wilson SH, Schlick T. J. Am. Chem. Soc. 2004 126, 8441–8453. [PubMed: 15238001]

[49]. Yang L, Beard WA, Wilson SH, Roux B, Broyde S, Schlick T. J. Mol. Biol. 2002 321, 459–478. [PubMed: 12162959]

[50]. Yang L, Beard WA, Wilson SH, Broyde S, Schlick T. J. Mol. Biol. 2002 317, 651–671. [PubMed: 11955015]

[51]. Yang Y, Qin Y, Ding Q, Bakhtina M, Wang L, Tsai M-D, Zhong D. Biochemistry 2014 53, 5405–5413. [PubMed: 25105470]

[52]. Kim T, Freudenthal BD, Beard WA, Wilson SH, Schlick T. Nucleic Acids Res. 2016 44, 4409–4424. [PubMed: 27034465]

[53]. Arora K, Beard WA, Wilson SH, Schlick T. Biochemistry 2005 44, 13328–13341. [PubMed: 16201758]

[54]. Rittenhouse RC, Apostoluk WK, Miller JH, Straatsma TP. Proteins: Struct., Funct., Bioinf. 2003 53, 667–682.

[55]. Xiang Y, Goodman MF, Beard WA, Wilson SH, Warshel A. Proteins: Struct., Funct., Bioinf. 2007 70, 231–247.

[56]. Ucisik MN, Hammes-Schiffer S. J. Chem. Inf. Model. 2015 55, 2672–2681. [PubMed: 26562587]

[57]. Foley MC, Schlick T. J. Am. Chem. Soc. 2008 130, 3967–3977. [PubMed: 18307346]

[58]. Su Z-Y, Lee W-J, Su W-S, Wang Y-T. Int. J. Mol. Sci. 2012 13, 7138–7148. [PubMed: 22837683]

[59]. Hu S, Cucinotta FA. J. Mol. Model. 2012 18, 1935–1949. [PubMed: 21870196]

[60]. Maffeo C, Luan B, Aksimentiev A. Nucleic Acids Res. 2012 40, 3812–3821. [PubMed: 22241779]

[61]. Olufsen M, Smalås AO, Brandsdal BO. J. Mol. Model. 2008 14, 201–213. [PubMed: 18196298]

[62]. Da L-T, Shi Y, Ning G, Yu J. Nucleic Acids Res. 2018 46, 568–581. [PubMed: 29253232]

[63]. Da L-T, Yu J. Nucleic Acids Res. 2018 46, 5410–5425. [PubMed: 29762710]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[64]. Bunta J, Dahlberg M, Eriksson L, Korolev N, Laaksonen A, Lohikoski R, Lyubartsev A, Pinak M, Schyman P. Int.J. Quantum Chem. 2006 107, 279–291.

[65]. Knips A, Zacharias M. Biopolymers 2014 103,215–222.

[66]. Tsutakawa SE, Yan C, Xu X, Weinacht CP, Freudenthal BD, Yang K, Zhuang Z, Washington MT, Tainer JA, Ivanov I. Structure 2015 23, 724–733. [PubMed: 25773143]

[67]. Querol-Audi J, Yan C, Xu X, Tsutakawa SE, Tsai M-S, Tainer JA, Cooper PK, Nogales E, Ivanov I. Proc. Natl. Acad. Sci. U. S. A. 2012 109, 8528–8533. [PubMed: 22586102]

[68]. Gattuso H, Durand E, Bignon E, Morell C, Georgakilas AG, Dumont E, Chipot C,Dehez F, Monari A. J. Phys. Chem. Lett. 2016 7, 3760–3765. [PubMed: 27612215]

[69]. Uckelmann M, Sixma TK. DNA Rep. 2017 56,92–101. Cutting-edge Perspectives in Genomic Maintenance IV.

[70]. Xu X, Yan C, Wohlhueter R, Ivanov I. Comput. Struct. Biotechnol. J. 2015 13,492–503. [PubMed: 26557958]

[71]. Mohan M, Pandya V, Anindya R. J. Mol. Graph. Model. 2018 84, 29–35. [PubMed: 29803030]

[72]. Mu H, Geacintov NE, Min J-H, Zhang Y, Broyde S. Chem. Res. Toxicol. 2017 30, 1344–1354. [PubMed: 28460163]

[73]. Mu H, Geacintov NE, Zhang Y, Broyde S. Biochemistry 2015 54, 5263–5267. [PubMed: 26270861]

[74]. Mu H, Kropachev K, Wang L, Zhang L, Kolbanovskiy A, Kolbanovskiy M, Geacintov NE, Broyde S. Nucleic Acids Res. 2012 40, 9675–9690. [PubMed: 22904073]

[75]. Qi Y, Spong MC, Nam K, Banerjee A, Banerjee A, Jiralerspong S, Karplus M, Verdine GL. Nature 2009 462, 762–766. [PubMed: 20010681]

[76]. Pieniazek SN, Hingorani MM, Beveridge D. Biophys.J. 2011 101, 1730–1739. [PubMed: 21961599]

[77]. Ishida H, Matsumoto A. Proteins: Struct., Funct., Bioinf. 2016 84, 1287–1303.

[78]. Sharma M, Predeus AV, Kovacs N, Feig M. Biophys.J. 2014 106, 2483–2492. [PubMed: 24896128]

[79]. Wang B, Francis J, Sharma M, Law SM, Predeus AV, Feig M. PLoS Comput. Biol. 2016 12, 1–23.

[80]. Salsbury FR Jr, Clodfelter JE, Gentry MB, Hollis T, Scarpinato KD. Nucleic Acids Res. 2006 34, 2173–2185. [PubMed: 16648361]

[81]. Mukherjee S, Law SM, Feig M. Biophys.J. 2009 96, 1707–1720. [PubMed: 19254532]

[82]. L. Negureanu FRS Jr. J. Biomol. Struct. Dyn. 2012 30, 347–361. [PubMed: 22712459]

[83]. Stofer E, Chipot C, Lavery R. J.Am. Chem. Soc. 1999 121, 9503–9508.

[84]. Nikolova EN, Bascom GD, Andricioaei I, Al-Hashimi HM. Biochemistry 2012 51, 8654–8664. [PubMed: 23035755]

[85]. Imhof P, Zahran M. PLoS One 2013 8, 1–11.

[86]. Banavali NK, MacKerell AD. J. Mol. Biol. 2002 319, 141–160. [PubMed: 12051942]

[87]. Song K, Campbell AJ, Bergonzo C, de los Santos C, Grollman AP, Simmerling C. J. Chem. Theory Comput. 2009 5,3105–3113. [PubMed: 26609990]

[88]. Yin Y, Yang L, Zheng G, Gu C, Yi C, He C, Gao YQ, Zhao XS. Proc. Natl. Acad. Sci. U. S.A. 2014 111, 8043–8048. [PubMed: 24843124]

[89]. Helabad MB, Kanaan N, Imhof P. Int.J. Mol. Sci. 2014 15, 11799–11816. [PubMed: 24995694]

[90]. Sharma M, Predeus AV, Mukherjee S, Feig M. J. Phys. Chem. B 2013 117, 6194–6205. [PubMed: 23621762]

[91]. Hu J, Ma A, Dinner AR. J. Chem. Phys. 2006 125, 114101. [PubMed: 16999460]

[92]. Kara M, Zacharias M. Biophys.J. 2013 104, 1089–1097. [PubMed: 23473492]

[93]. Curuksu J, Zacharias M, Lavery R, Zakrzewska K. Nucleic Acids Res. 2009 37,3766–3773. [PubMed: 19380377]

[94]. La Rosa G, Zacharias M. Nucleic Acids Res. 2016 44, gkw827.

[95]. Shvetsov AV, Lebedev DV, Chervyakova DB, Bakhlanova IV, Yung IA, Radulescu A, Kuklin AI, Baitin DM, Isaev-Ivanov VV. FEBS Lett. 2014 588, 948–955. [PubMed: 24530684]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[96]. Kim R, Kanamaru S, Mikawa T, Prévost C, Ishii K, Ito K, Uchiyama S, Oda M, Iwasaki H, Kim SK, Takahashi M. Nucleic Acids Res. 2018 46, 2548–2559. [PubMed: 29390145]

[97]. Kokabu Y, Ikeguchi M. Biophys.J. 2013 104, 1556–1565. [PubMed: 23561532]

[98]. Buis N, Skylaris C-K, Grant GH, Rajendra E, Payne MC, Venkitaraman AR. Mol. Sim. 2008 34, 749–759.

[99]. Kang H, Yoo J, Sohn B-K, Lee S-W, Lee HS, Ma W, Kee J-M, Aksimentiev A, Kim H. Nucleic Acids Res. 2018 46, 9401–9413. [PubMed: 30032232]

[100]. Yoo J, Kim H, Aksimentiev A, Ha T. Nat. Commun. 2016 7, 11045. [PubMed: 27001929]

[101]. Yu J, Ha T, Schulten K. Nucleic Acids Res. 2004 32, 6683–6695. [PubMed: 15613597]

[102]. Wheatley EG, Pieniazek SN, Mukerji I, Beveridge D. Biophys.J. 2012 102, 552–560. [PubMed: 22325278]

[103]. Karymov MA, Chinnaraj M, Bogdanov A, Srinivasan AR, Zheng G, Olson WK, Lyubchenko YL. Biophys.J. 2008 95,4372–4383. [PubMed: 18658216]

[104]. Yoo J, Aksimentiev A. Phys. Chem. Chem. Phys. 2018 20, 8432–8449. [PubMed: 29547221]

[105]. Priyakumar UD, MacKerell AD. Chem. Rev. 2006 106, 489–505. [PubMed: 16464016]

[106]. Mura C, McCammon JA. Nucleic Acids Res. 2008 36, 4941–4955. [PubMed: 18653524]

[107]. Kara M, Drsata T, Lankas F, Zacharias M. Biopolymers 2015 103, 23–32. [PubMed: 25130987]

[108]. Meli M, Sustarsic M, Craggs TD, Kapanidis AN, Colombo G. Front. Mol. Biosci. 2016 3, 20. [PubMed: 27303671]

[109]. Radhakrishnan R, Schlick T. Proc. Natl. Acad. Sci. U. S. A. 2004 101, 5970–5975. [PubMed: 15069184]

[110]. Ma W, Whitley KD, Chemla YR, Luthey-Schulten Z, Schulten K. Elife 2018 7, e34186. [PubMed: 29664402]

[111]. Mhashal AR, Choudhury CK, Roy S. J. Mol. Model. 2016 22, 54. [PubMed: 26860503]

[112]. Liebl K, Zacharias M. J. Phys. Chem. B 2017 121, 11019–11030. [PubMed: 29064703]

[113]. Knips A, Zacharias M. Sci. Reports 2017 7,41324.

[114]. Radhakrishnan R, Schlick T. J.Am. Chem. Soc. 2005 127, 13245–13252. [PubMed: 16173754]

[115]. Radhakrishnan R, Arora K, Wang Y, Beard WA, Wilson SH, Schlick T. Biochemistry 2006 45, 15142–15156. [PubMed: 17176036]

[116]. Li Y, Freudenthal BD, Beard WA, Wilson SH, Schlick T. J. Am. Chem. Soc. 2014 136, 3630–3639. [PubMed: 24511902]

[117]. Li Y, Gridley CL, Jaeger J, Sweasy JB, Schlick T. J. Am. Chem. Soc. 2012 134, 9999–10010. [PubMed: 22651551]

[118]. Ishida H. J. Comput. Chem. 2010 31, 2317–2329. [PubMed: 20575014]

[119]. Xiang Y, Oelschlaeger P, Florián J, Goodman MF, Warshel A. Biochemistry 2006 45, 7036–7048. [PubMed: 16752894]

[120]. Florián J, Goodman MF, Warshel A. Biopolymers 2003 68, 286–299. [PubMed: 12601790]

[121]. Bren U, Martinek V, Florián J. J. Phys. Chem. B 2006 110, 10557–10566. [PubMed: 16722767]

[122]. Florián J, Goodman MF, Warshel A. J.Am. Chem. Soc. 2003 125, 8163–8177. [PubMed: 12837086]

[123]. Klvarňa M, Bren U, Florián J. J. Phys. Chem. B 2016 120, 13017–13030. [PubMed: 27992186]

[124]. Florián J, Warshel A, Goodman MF. J. Phys. Chem. B 2002 106, 5754–5760.

[125]. Rucker R, Oelschlaeger P, Warshel A. Proteins: Struct., Funct., Bioinf 2009 78, 671–680.

[126]. Hu S, Cucinotta FA. Radiat. Prot. Dosimetry 2011 143, 196–201. [PubMed: 21196465]

[127]. Rousseau BJG, Shafei S, Migliore A, Stanley RJ, Beratan DN. J.Am. Chem. Soc. 2018 140, 2853–2861. [PubMed: 29401372]

[128]. Zhong S, Chen X, Zhu X, Dziegielewska B, Bachman KE, Ellenberger T, Ballin JD, Wilson GM, Tomkinson AE, MacKerell AD. J. Med. Chem. 2008 51, 4553–4562. [PubMed: 18630893]

[129]. Lin P, Pedersen LC, Batra VK, Beard WA, Wilson SH, Pedersen LG. Proc. Natl. Acad. Sci. U. S.A. 2006 103, 13294–13299. [PubMed: 16938895]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[130]. Stevens DR, Hammes-Schiffer S. J.Am. Chem. Soc. 2018 140, 8965–8969. [PubMed: 29932331]

[131]. Rosta E, Woodcock HL, Brooks BR, Hummer G. J. Comput. Chem. 2009 30, 1634–1641. [PubMed: 19462398]

[132]. Sgrignani J, Magistrato A. ACS Catal. 2015 5, 3864–3875.

[133]. Rosta E, Nowotny M, Yang W, Hummer G. J. Am. Chem. Soc. 2011 133, 8934–8941. [PubMed: 21539371]

[134]. Wang Y, Schlick T. J. Am. Chem. Soc. 2008 130, 13240–13250. [PubMed: 18785738]

[135]. Alberts IL, Wang Y, Schlick T. J. Am. Chem. Soc. 2007 129, 11100–11110. [PubMed: 17696533]

[136]. Radhakrishnan R, Schlick T. Biochem. Biophys. Res. Commun. 2006 350, 521–529. [PubMed: 17022941]

[137]. Perera L, Freudenthal BD, Beard WA, Pedersen LG, Wilson SH. Nucleic Acids Res. 2017 45, 2736–2745. [PubMed: 28108654]

[138]. Bojin MD, Schlick T. J. Phys. Chem. B 2007 111, 11244–11252. [PubMed: 17764165]

[139]. Cisneros GA, Perera L, Garcia-Diaz M, Bebenek K, Kunkel TA, Pedersen LG. DNA Rep. 2008 7, 1824–1834.

[140]. Zheng X, Garcia J, Stuchebrukhov AA. J. Phys. Chem. B 2008 112, 8724–8729. [PubMed: 18588340]

[141]. Sadeghian K, Flaig D, Blank ID, Schneider S, Strasser R, Stathis D, Winnacker M, Carell T, Ochsenfeld C. Angew. Chem. Int. Ed. 2014 53, 10044–10048.

[142]. Poater J, Swart M, Guerra CF, Bickelhaupt FM. Chem. Commun. 2011 47, 7326–7328.

[143]. Brovarets OO', Zhurakivsky RO, Hovorun DM. J. Biomol. Struct. Dyn. 2015 33, 674–689. [PubMed: 24650179]

[144]. Brovarets OO', Hovorun DM. J. Biomol. Struct. Dyn. 2015 33, 925–945. [PubMed: 24842163]

[145]. van Bochove MA, Roos G, Fonseca Guerra C, Hamlin TA, Bickelhaupt FM. Chem. Commun. 2018 54,3448–3451.

[146]. Zhang L, Li W, Fang T, Li S. Theor. Chem. Acc 2016 135, 34.

[147]. Fonseca Guerra C, Bickelhaupt FM, Baerends EJ. Cryst. Growth Des. 2002 2, 239–245.

[148]. Poater J, Swart M, Guerra CF, Bickelhaupt FM. Comput. Theor. Chem. 2012 998, 57–63.

[149]. Poater J, Swart M, Bickelhaupt FM, Fonseca Guerra C. Org. & Biomol. Chem. 2014 12, 4691–4700. [PubMed: 24871817]

[150]. Hamlin TA, Poater J, Fonseca Guerra C, Bickelhaupt FM. Phys. Chem. Chem. Phys. 2017 19, 16969–16978. [PubMed: 28497822]

[151]. Florián J, Goodman MF, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2005 102, 6819–6824. [PubMed: 15863620]

[152]. Adelman JL, Chodera JD, Kuo I-FW, Miller TF, Barsky D. Biophys. J. 2010 98, 3062–3069. [PubMed: 20550919]

[153]. Daitchman D, Greenblatt HM, Levy Y. Nucleic Acids Res. 2018 46, 5935–5949. [PubMed: 29860305]

[154]. Chen R, Subramanyam S, Elcock AH, Spies M, Wold MS. Nucleic Acids Res. 2016 44, 5758–5772. [PubMed: 27131385]

[155]. Shimizu M, Noguchi Y, Sakiyama Y, Kawakami H, Katayama T, Takada S. Proc. Natl. Acad. Sci. U. S.A. 2016 113, E8021–E8030. [PubMed: 27911788]

[156]. Mishra G, Levy Y. Proc. Natl. Acad. Sci. U. S. A. 2015 112, 5033–5038. [PubMed: 25855635]

[157]. Liu H, Shi Y, Chen XS, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2009 106, 7449–7454. [PubMed: 19383795]

[158]. Yoshimoto K, Arora K, Brooks CL. Biophys. J. 2010 98, 1449–1457. [PubMed: 20409463]

[159]. Wang W, Nocka LM, Wiemann BZ, Hinckley DM, Mukerji I, Starr FW. Sci. Reports 2016 6, 22863.

[160]. van Buuren BNM, Hermann T, Wijmenga SS, Westhof E. Nucleic Acids Res. 2002 30, 507–514. [PubMed: 11788713]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[161]. Lee JY, Terakawa T, Qi Z, Steinfeld JB, Redding S, Kwon Y, Gaines WA, Zhao W, Sung P, Greene EC. Science 2015 349, 977–981. [PubMed: 26315438]

[162]. Li W, Nordenskiöld L, Mu Y. J. Phys. Chem. B 2011 115, 14713–20. [PubMed: 22035057]

[163]. Ando T, Skolnick J. PLoS Comput. Biol. 2014 10, 1–10.

[164]. Guardiani C, Cencini M, Cecconi F. Phys. Bio. 2014 11, 026003. [PubMed: 24685517]

[165]. Bhattacherjee A, Krepel D, Levy Y. WIREs Comput Mol Sci 2016 6,515–531.

[166]. Saladin A, Amourda C, Poulain P, Férey N, Baaden M, Zacharias M, Delalande O, Prévost C. Nucleic Acids Res. 2010 38, 6313–6323. [PubMed: 20507912]

[167]. Setny P, Bahadur RP, Zacharias M. BMC Bioinformatics 2012 13, 228. [PubMed: 22966980]

[168]. Gauthier MG, Norio P, Bechhoefer J. PLoS One 2012 7, 1–13.

[169]. Gauthier MG, Herrick J, Bechhoefer J. Phys. Rev. Lett. 2010 104, 218104. [PubMed: 20867141]

[170]. Cucinotta FA, Pluth JM, Anderson JA, Harper JV, O'Neill P. Radiat. Res. 2008 169, 214–222. [PubMed: 18220463]

[171]. Taleei R, Nikjoo H. Radiat. Res. 2013 179, 530–539. [PubMed: 23560635]

[172]. Dorfman KD, Fulconis R, Dutreix M, Viovy J-L. Phys. Rev. Lett. 2004 93, 268102. [PubMed: 15698024]

[173]. Kochugaeva AAK, Maria P Shvets AB Biophys.J. 2017 112, 859–867. [PubMed: 28297645]

[174]. Pierobon P, Miné-Hattab J, Cappello G, Viovy J-L, Lagomarsino MC. Phys. Rev. E. 2010 82, 061904.

[175]. van der Heijden T, Seidel R, Modesti M, Kanaar R, Wyman C, Dekker C. Nucleic Acids Res. 2007 35, 5646–5657. [PubMed: 17709342]

[176]. de Moura APS, Retkute R, Hawkins M, Nieduszynski CA. Nucleic Acids Res. 2010 38, 5623–5633. [PubMed: 20457753]

[177]. Goldar A, Labit H, Marheineke K, Hyrien O. PLoS One 2008 3, 1–8.

[178]. Lygeros J, Koutroumpas K, Dimopoulos S, Legouras I, Kouretas P, Heichinger C, Nurse P, Lygerou Z. Proc. Natl. Acad. Sci. U. S. A. 2008.

[179]. Hyrien O, Goldar A. Chromosome Res. 2010 18, 147–161. [PubMed: 20205354]

[180]. Gispan A, Carmi M, Barkai N. Genome Res. 2017 27, 310–319. [PubMed: 28028072]

[181]. Retkute R, Nieduszynski CA, de Moura A. Phys. Rev. E 2012 86, 031916.

[182]. Bechhoefer J, Rhind N. Trends Gen. 2012 28,374–381.

[183]. Fok P-W, Guo C-L, Chou T. J. Chem. Phys. 2008 129, 235101. [PubMed: 19102557]

[184]. Semenenko VA, Stewart RD, Ackerman EJ. Radiat. Res. 2005 164, 180–193. [PubMed: 16038589]

[185]. Semenenko VA, Stewart RD. Radiat. Res. 2005 164, 194–201. [PubMed: 16038590]

[186]. Friedland W, Dingfelder M, Kundrät P, Jacob P. Mutat. Res. 2011 711, 28–40. [PubMed: 21281649]

[187]. Dolan D, Nelson G, Zupanic A, Smith G, Shanley D. PLoS One 2013 8, e55190. [PubMed: 23457464]

[188]. Fulconis R, Dutreix M, Viovy J-L. Biophys. J. 2005 88 6, 3770–3779. [PubMed: 15749781]

[189]. Ma CJ, Gibb B, Kwon Y, Sung P, Greene EC. Nucleic Acids Res. 2017 45, 749–761. [PubMed: 27903895]

[190]. Alder BJ, Wainwright TEW. J. Chem. Phys. 1959 31, 459–466.

[191]. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J. Comput. Chem. 2005 26, 1781–1802. [PubMed: 16222654]

[192]. Perilla JR, Schulten K. Nat. Commun. 2017 8, 15959. [PubMed: 28722007]

[193]. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J. Am. Chem. Soc. 1995 117, 5179–5197.

[194]. Foloppe N, MacKerell AD Jr. J. Comput. Chem. 2000 21, 86–104.

[195]. Jorgensen WL, Maxwell DS, Tirado-Rives J. J. Am. Chem. Soc. 1996 118, 11225–11236.

[196]. Daura X, Mark AE, van Gunsteren WF. J. Comput. Chem. 1998 19, 535–547.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[197]. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. SoftwareX 2015 1, 19–25.

[198]. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. Comput. Phys. Commun. 1995 91, 1–41.

[199]. Brooks BR, Brooks CL, MacKerell AD Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. J. Comput. Chem. 2009 3G, 1545–614.

[200]. Hünenberger PH. Thermostat Algorithms for Molecular Dynamics Simulations, pages 105–149. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[201]. Koopman EA, Lowe CP. J. Chem. Phys. 2006 124, 204103. [PubMed: 16774315]

[202]. Nosé S, Klein ML. Mol. Phys. 1983 5G, 1055–76.

[203]. Nosé S. J. Chem. Phys. 1984 81, 511–519.

[204]. Hoover WG. Phys. Rev. A 1985 31, 1695–1697.

[205]. Martyna GJ, Tobias DJ, Klein ML. J. Chem. Phys. 1994 101, 4177–4189.

[206]. Stoyanov SD, Groot RD. Journal of Chemical Physics 2005 122, 14112. [PubMed: 15638647]

[207]. Noid WG. J. Chem. Phys. 2013 139, 090901. [PubMed: 24028092]

[208]. Soper A. Chem. Phys. 1996 202, 295–306.

[209]. Lyubartsev AP, Laaksonen A. Phys. Rev. E 1995 52, 3730–3737.

[210]. Lyubartsev AP, Laaksonen A. Phys. Rev. E 1997 55, 5689–5696.

[211]. Izvekov S, Voth GA. J. Chem. Phys. 2005 123, 134105. [PubMed: 16223273]

[212]. Noid WG, Chu J-W, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC. J. Chem. Phys. 2008 128, 244114. [PubMed: 18601324]

[213]. Marrink SJ, Tieleman DP. Chem. Soc. Rev. 2013 42, 6801–6822. [PubMed: 23708257]

[214]. Maffeo C, Ngo TTM, Ha T, Aksimentiev A. J. Chem. Theory Comput. 2014 10, 2891–2896. [PubMed: 25136266]

[215]. Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, Marrink SJ. WIREs Comput Mol Sci 2014 4,225–248.

[216]. Zwanzig R. Proc. Natl. Acad. Sci. U. S. A. 1988 85, 2029–2030. [PubMed: 3353365]

[217]. Warshel A, Levitt M. J. Mol. Biol. 1976 103, 227–249. [PubMed: 985660]

[218]. Senn HM, Thiel W. Angew. Chem. Int. Ed. 2009 48, 1198–1229.

[219]. Vreven T, Morokuma K, Farkas Ö, Schlegel HB, Frisch MJ. J. Comput. Chem. 2003 24, 760–769. [PubMed: 12666168]

[220]. Cao L, Ryde U. Front. Chem. 2018 6, 89. [PubMed: 29666794]

[221]. Kamerlin SCL, Warshel A. WIREs Comput Mol Sci 2011 1,30–45.

[222]. Warshel A, Weiss RM. J. Am. Chem. Soc. 1980 102, 6218.

[223]. Hwang JK, Chu ZT, Yadav A, Warshel A. J. Phys. Chem. 1991 95, 8445–8448.

[224]. Aquist J, Warshel A. Chem. Rev. 1993 93, 2523–2544.

[225]. Schlegel HB, Sonnenberg JL. J. Chem. Theory Comput. 2006 2, 905–911. [PubMed: 26633049]

[226]. Warshel A. Annu. Rev. Biophys. Biomol. Struct. 2003 32, 425–443. [PubMed: 12574064]

[227]. Zwanzig RW. J. Chem. Phys. 1954 22, 1420–1426.

[228]. Jorgensen WL, Thomas LL. J. Chem. Theory Comput. 2008 4, 869–876. [PubMed: 19936324]

[229]. Åqvist J, Warshel A. J. Am. Chem. Soc. 1990 112, 2860–2868.

[230]. Lee FS, Chu Z-T, Bolger MB, Warshel A. Protein Eng. Des. Sel. 1992 5, 215–228.

[231]. Lee FS, Chu ZT, Warshel A. J. Comput. Chem. 1993 14, 161–185.

[232]. Sham YY, Chu ZT, Tao H, Warshel A. Proteins: Struct., Funct., Bioinf. 2000 39,393–407.

[233]. Singh N, Warshel A. Proteins: Struct., Funct., Bioinf. 2010 78, 1705–1723.

[234]. Torrie GM, Valleau JP. Chem. Phys. Lett. 1974 28, 578–581.

[235]. Torrie GM, Valleau JP. J. Comput. Phys. 1977 23, 187–199.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[236]. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. J. Comput. Chem. 1992 13, 1011–1021.

[237]. Souaille M, Roux B. Comput. Phys. Commun. 2001 135, 40–57.

[238]. Roux B. Comput. Phys. Commun. 1995 91, 275–282.

[239]. Kästner J. WIREs Comput Mol Sci 2011 1,932–942.

[240]. Darve E, Pohorille A. J. Chem. Phys. 2001 115, 9169–9183.

[241]. Darve E, Wilson D, Pohorille A. Mol. Sim. 2002 28, 113–144.

[242]. Comer J, Gumbart JC, Henin J, Lelièvre T, Pohorille A, Chipot C. J. Phys. Chem. B 2015 119, 1129–1151. [PubMed: 25247823]

[243]. Rodriguez-Gomez D, Darve E, Pohorille A. J. Chem. Phys. 2004 120, 3563–3578. [PubMed: 15268518]

[244]. Dellago C, Bolhuis PG, Csajka FS, Chandler D. J. Chem. Phys. 1998 108, 1964–1977.

[245]. Dellago C, Bolhuis PG, Chandler D. J. Chem. Phys. 1998 108, 9236[270] 9245.

[246]. Pratt LR. J. Chem. Phys. 1986 85, 5045–5048.

[247]. Bolhuis PG, Dellago C, Chandler D. Faraday Discuss. 1998 110, 421–436.

[248]. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Annu. Rev. Phys. Chem. 2002 53, 291–318. [PubMed: 11972010]

[249]. Dellago C, Bolhuis PG, Geissler PL. In Prigogine I, Rice SA, editors, Adv. Chem. Phys, chapter 1, pages 1–78. John Wiley & Sons, Inc., New York, 2003.

[250]. Dellago C, Bolhuis PG. In Reiher M, editor, Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations, pages 291–317. Springer, Berlin, Heidelberg, 2007.

[251]. Yang L, Qin Gao Y.. J. Chem. Phys. 2009 131, 214109. [PubMed: 19968339]

[252]. Pande VS, Beauchamp K, Bowman GR. Meth. 2010 52, 99–105. Protein Folding.

[253]. Langston LD, O'Donnell M. Mol. Cell 2006 23, 155–160. [PubMed: 16857582]

[254]. Zhou R, Kozlov AG, Roy R, Zhang J, Korolev S, Lohman TM, Ha T. Cell 2011 146, 222–232. [PubMed: 21784244]

[255]. Moldovan G-L, Pfander B, Jentsch S. Cell2007 129, 665–79. [PubMed: 17512402]

[256]. Indiani C, McInerney P, Georgescu R, Goodman MF, O'Donnell M. Mol. Cell 2005 19, 805–815. [PubMed: 16168375]

[257]. Holt IJ, Reyes A. Cold Spring Harb. Perspect. Biol. 2012.

[258]. Hanahan D, Weinberg RA. Cell 2011 144, 646–674. [PubMed: 21376230]

[259]. Gentile F, Tuszynski JA, Barakat KH. Curr. Pharm. Des. 2016 22, 3527–3546. [PubMed: 27093954]

[260]. Sun J, Lee K-J, Davis AJ, Chen DJ. J. Biol. Chem. 2012 287, 4936–4945. [PubMed: 22179609]

[261]. Monti MC, Cohen SX, Fish A, Winterwerp HHK, Barendregt A, Friedhoff P, Perrakis A, Heck AJR, Sixma TK, van den Heuvel RHH, Lebbink JHG. Nucleic Acids Res. 2011 39,8052–8064. [PubMed: 21737427]

[262]. Schärer OD. Cold Spring Harb. Perspect. Biol. 2013 5, a012609. [PubMed: 24086042]

[263]. Min J-H, Pavletich NP. Nature 2007 449, 570–575. [PubMed: 17882165]

[264]. Conlin MP, Reid DA, Small GW, Chang HH, Watanabe G, Lieber MR, Ramsden DA, Rothenberg E. Cell Reports 2017 20,2810–2819. [PubMed: 28930678]

[265]. Branzei D, Foiani M. Nat. Rev. Mol. Cell Biol. 2008 9, 297–308. [PubMed: 18285803]

[266]. Cox MM. Annu. Rev. Microbiol. 2003 57, 551–77. [PubMed: 14527291]

[267]. Cox MM. Nat. Rev Mol. Cell Biol. 2007 8, 127–38. [PubMed: 17228330]

[268]. Chen Z, Yang H, Pavletich NP. Nature 2008 453, 489–494. [PubMed: 18497818]

[269]. Kornyshev AA, Leikin S. Phys. Rev. Lett. 2001 86, 3666–3669. [PubMed: 11328049]

[270]. Baldwin GS, Brooks NJ, Robson RE, Wynveen A, Goldar A, Leikin S, Seddon JM, Kornyshev AA. J. Phys. Chem. B 2008 112, 1060–1064. [PubMed: 18181611]

[271]. Li L, Pabit SA, Meisburger SP, Pollack L. Phys. Rev Lett. 2011 106, 108101. [PubMed: 21469837]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[272]. Drees JC, Lusetti SL, Chitteni-Pattu S, Inman RB, Cox MM. Mol. Cell 2004 15, 789–798. [PubMed: 15350222]

[273]. Duderstadt KE, Reyes-Lamothe R, van Oijen AM, Sherratt DJ. Cold Spring Harb. Perspect. Biol. 2014 6, a010157. [PubMed: 23881939]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

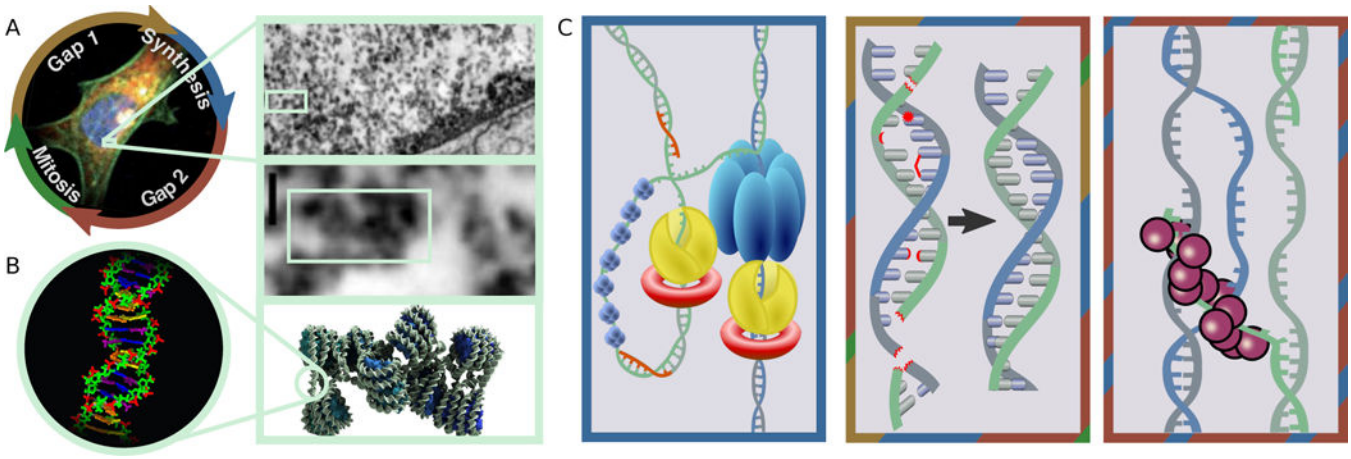


Figure 1:

Genome maintenance throughout the cell cycle is performed at the molecular scale. (A) The eukaryotic cell cycle. When a cell is first formed, it enters the Gap 1 phase (gold arrow). During the Synthesis phase (blue), the cell replicates its DNA. The cell prepares to divide during the Gap 2 phase (red). Finally, the cell's chromatin condenses into chromosomes, and mitosis occurs (green). The image of the fluorescently-labeled epithelial cell with blue-stained nucleus is reproduced with permission [6]. Copyright 2010, The Royal Society of Chemistry. (B) DNA provides the molecular basis for storing information needed by the cell. The three images on the right depict the connection from cell to the double helix through multiple scales. The top two panels depict electron microscopy images of chromatin reproduced with permission [7]. Copyright 2017, The American Association for the Advancement of Science. The bottom image depicts a computational molecular model of a nucleosome array—the basis of chromatin. The computational modeling provides detail at length scales that are challenging to resolve through experimental methods. (C) Molecular-scale depictions of genome processing. Each panel depicts the main genome-processing activities of the cell from a microscopic view. From left to right, the cell replicates, repairs, and (sometimes) recombines its genetic material. The color of the outline of each panel indicates the phase of the cell cycle in which the process occurs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

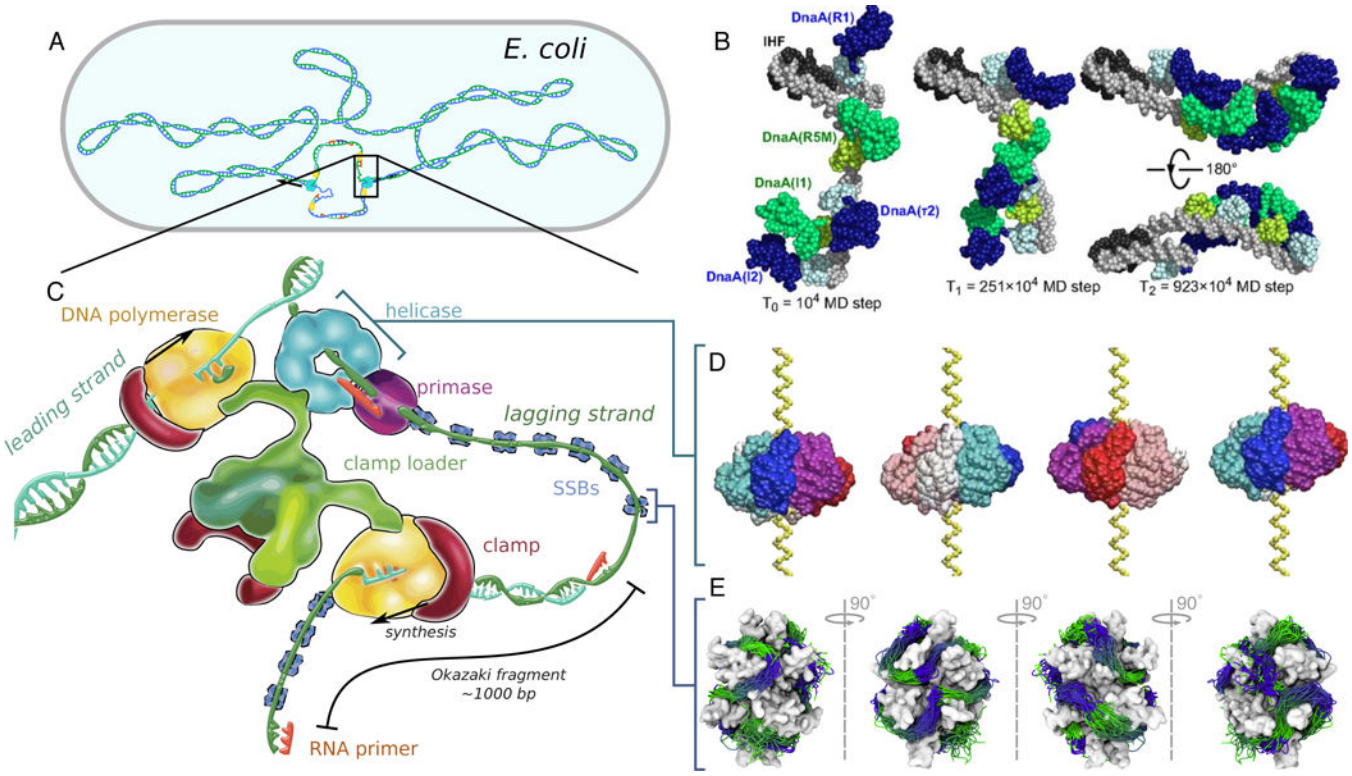


Figure 2:

Simulations of DNA replication systems. (A-C) DNA replication in model organism *Escherichia coli.* (A) Replication begins with the melting of the origin of replication, whereafter a protein complex known as a replication fork or replisome assembles at either end of the DNA bubble and begins DNA synthesis. (B) Oligomerization of proteins at the origin of replication. Initiation of DNA replication occurs by the protein-mediated melting of DNA at the origin of replication. Melting is driven in a mechanical process by the oligomerization of DnaA proteins in bacteria (ORC in eukaryotes). Coarse-grained simulations by the Takada group showed how DnaA proteins bound to specific sites at the origin interact to form a helix that puts mechanical strain on the DNA. Adapted with permission [155]. Copyright 2016, National Academy of Sciences. (C) The replication fork of the model organism *Escherichia coli.* Each replication fork consists of a hexameric ring-shaped helicase (light blue) that hydrolyses ATP to unwind the double-stranded DNA, driving the complex away from the origin of replication. A protein called primase (purple) begins replication on each of the daughter strands by synthesizing a short RNA strand or primer (orange) that is subsequently extended by a DNA polymerase (yellow) that synthesizes DNA in the 5'-to-3' direction. A ring-shaped "clamp" protein (red) circumscribes the DNA and provides a binding site for the DNA polymerase. A central clamp-loader protein complex (green) directly binds the helicase and polymerase enzymes and places a clamp at each RNA primer. Finally, single-stranded DNA binding proteins (blue) coat all exposed single-stranded DNA to prevent the formation of secondary structures that would prevent replication. Adapted with permission [253]. Copyright 2006, Elsevier. (D) Translocation of a hexameric helicase along DNA. The Brooks group demonstrated through coarse-grained modeling that sequential conformational changes in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

adjacent hexameric subunits unidirectionally drive the translocation of a helicase. Adapted with permission [158]. Copyright 2010, Elsevier. (E) All-atom simulations of single-stranded DNA binding protein (SSB) detail the extent of fluctuations of DNA bound to the protein surface. The simulations revealed the formation and resolution of small ssDNA bulges that are believed to provide a mechanism for SSB diffusion along DNA. Adapted with permission [45]. Oxford University Press.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

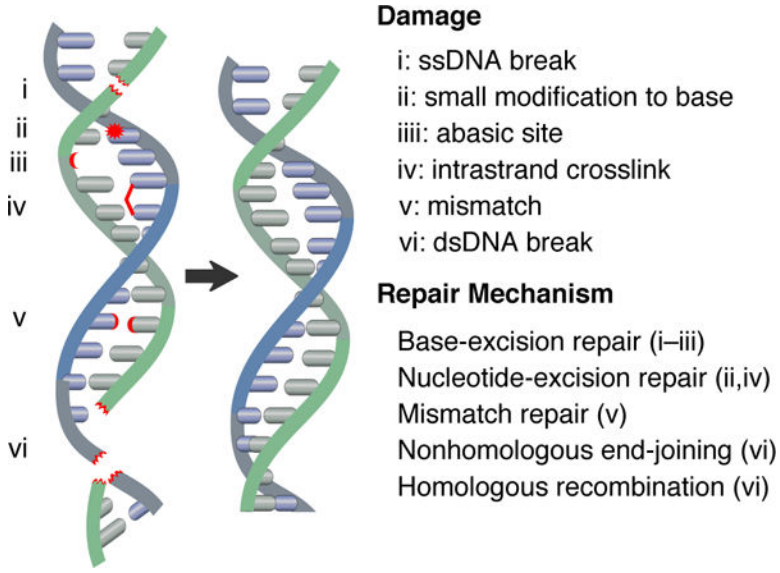


Figure 3:

Overview of DNA damage and repair mechanisms. Damage only to one strand is usually repaired through the mismatch, base-excision and nucleotide-excision repair mechanisms. Double stranded DNA breaks are repaired through the non-homologous end-joining pathway during most of the cell cycle when a sister chromatid is unavailable. After the sister chromatid has been synthesized and before mitosis (chromosome separation), such dsDNA breaks are repaired through homologous recombination.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

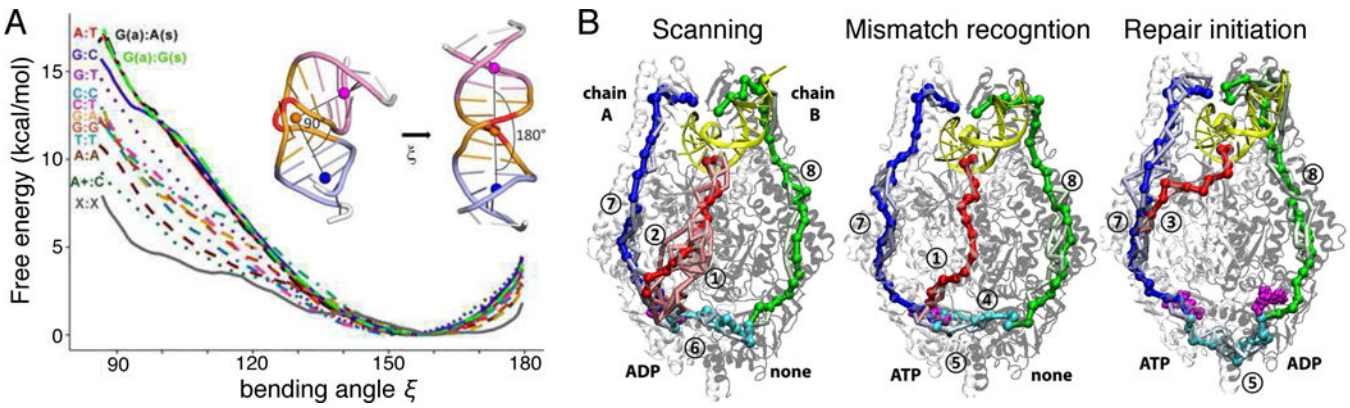


Figure 4:

Detection of DNA damage in mismatch repair. (A) Free energy cost of bending normal and mismatch-containing duplexes extracted from all-atom enhanced sampling simulations. The cost of bending duplexes is reduced by the presence of a mismatch. Adapted with permission [90]. Copyright 2013, American Chemical Society. (B) Allosteric pathways model of damage recognition by MutS according to the Feig group's study [79]. MutS (gray) is in a DNA-scanning state bound to a helix (yellow) with one ADP bound (magenta; left panel). If an unusually flexible helical region is encountered, allosteric communication (pathway shown in red) from the mismatch recognition domain to the ATP-binding domain allows for exchange of the bound ADP for an ATP molecule, increasing the allosteric coupling between the domains (center panel). If the helix is indeed damaged, allosteric changes allow a second ATP molecule to bind, which is immediately hydrolyzed to form ADP (right panel). The ATP binding domain now acts as a signal for other proteins to come and repair the DNA. The circled numbers correspond to particular allosteric communication pathways identified by the Feig group. Adapted under the terms of the CC BY 4.0 license [79]. Copyright 2016, Wang *et al.*

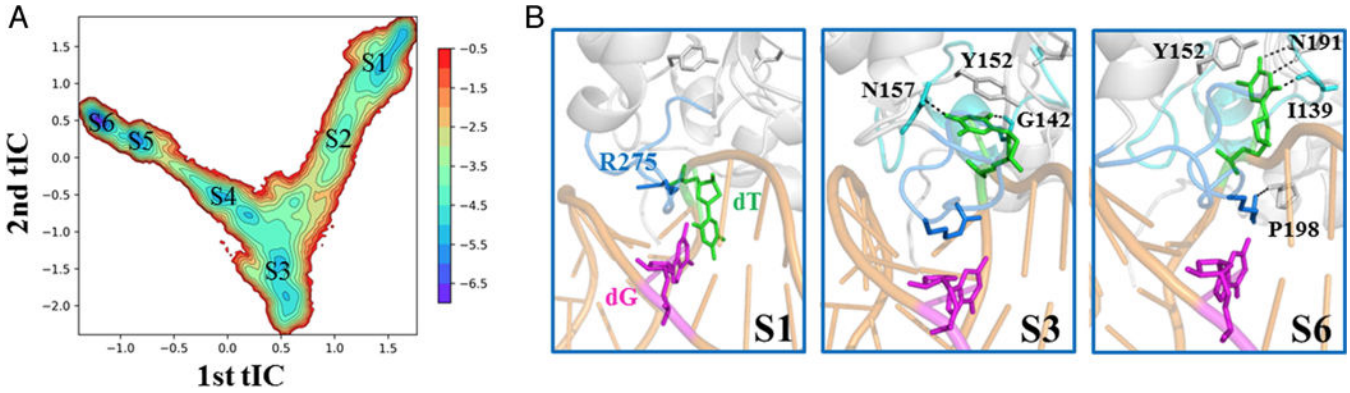


Figure 5:

All-atom MD simulations of damage recognition in base excision repair. (A) The free energy landscape for flipping a mismatched base out of a duplex. Over 30 μs of simulations of a glycosylase protein bo und to a helix carrying a damaged base allowed the construction of a Markov state model capable of describing the kinetics of DNA damage recognition by the protein. (B) Snapshots depicting the conformation of the protein-DNA complex in the most heavily populated states. After the base is flipped out of the helix and into the binding protein pocket, it is cleaved. Adapted with permission [63]. Copyright 2018, Oxford University Press.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

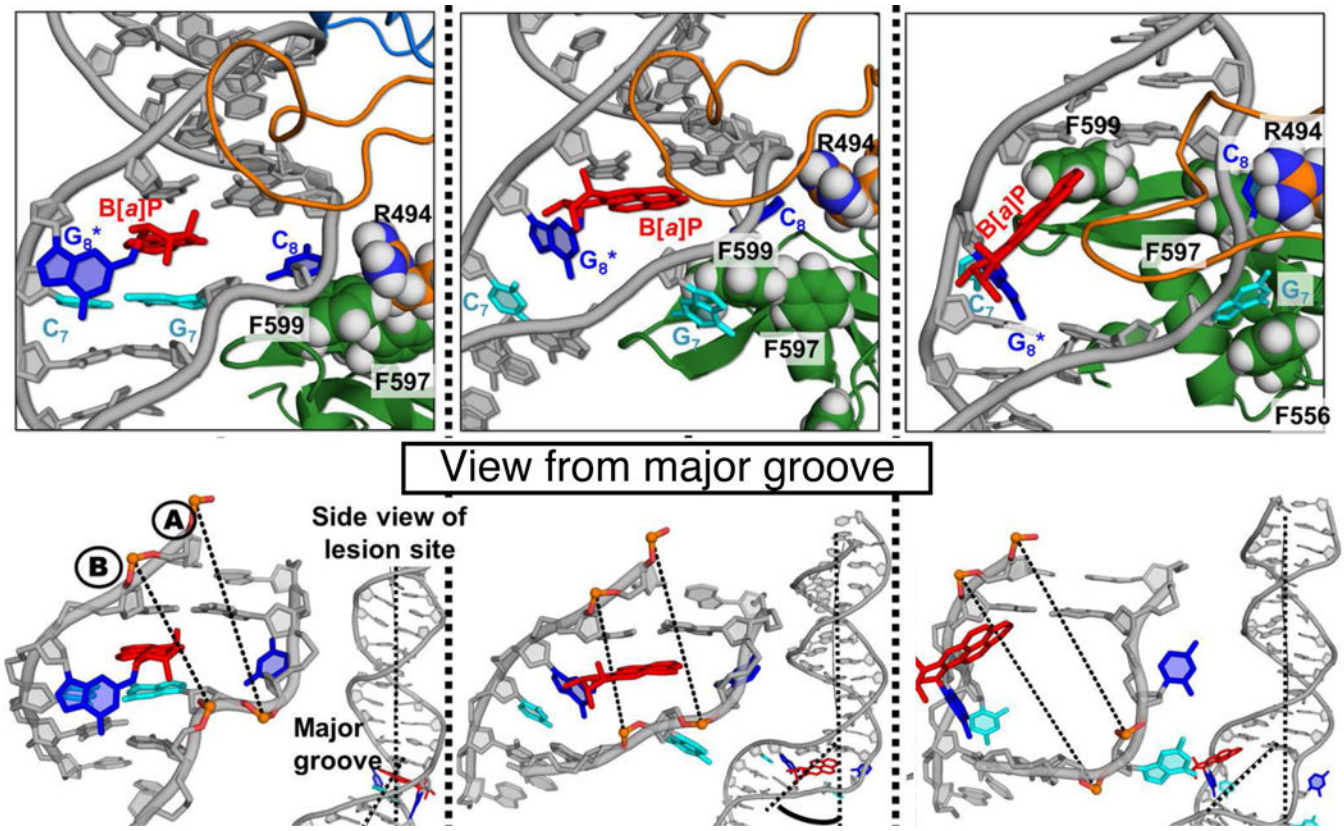


Figure 6:

All-atom MD simulation of damage recognition in nucleotide excision repair. (Left) The Rad4 protein (green) is in a damage-scanning conformation bound to a DNA helix (grey) distorted by damage (red). (Center) The protein interacts with a base on the undamaged strand, pulling it out of the helix. (Right) The protein pulls a second base of the helix, inserting phenylalanine residues to stabilize the helix and forcing the damaged base into solution where it can be accessed and repaired by other proteins. The snapshots in the bottom row omit the Rad4 protein to clearly show the distortions of the DNA duplex. Adapted with permission [72]. Copyright 2017, American Chemical Society.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

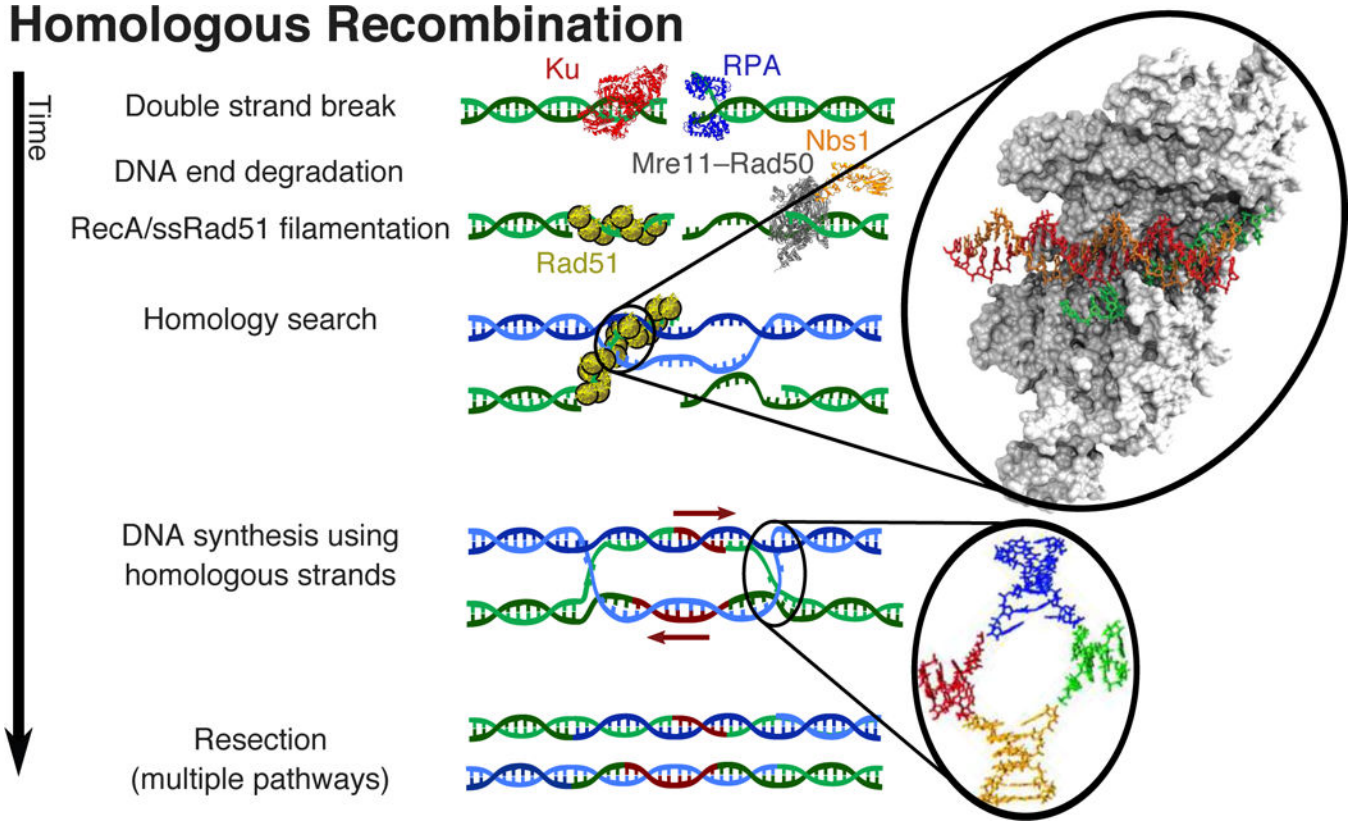


Figure 7:

Overview of the process of homologous recombination in eukaryotic cells. The major steps during recombination are shown from top to bottom. Homologous recombination begins with the detection of a double-stranded DNA break by Ku (red) and possibly RPA (blue; SSB in bacteria) proteins. The Mre11–Rad50–Nbs1 (MRN) complexes (gray and orange; RecBCD complex in bacteria) will form on each end of the break, cleaving the end and degrading one of the strands. Next, RPA coats the exposed single-stranded DNA recruiting Rad51 proteins (yellow; RecA in bacteria) that form a filament on the DNA to displace the RPA. The DNA bases are exposed in the Rad51–ssDNA filament, allowing a search for homologous DNA. The blown up region depicts a model of the filament–dsDNA encounter complex, adapted with permission [166]. Copyright 2010, Oxford University Press. Once the homologous region of a duplex is found, the Rad51 filament facilitates exchange between the strands. Holliday junctions are formed at either end of the break and polymerases (usually Pol δ or Pol I) synthesize the missing DNA, using the intact homologous DNA as a template. The Holliday junctions perform branch migration, diffusing randomly and promoting variation in the amount of DNA swapped between the sister chromatids. The blown up region depicts a model of a Holliday junction in its open configuration, reproduced with permission [101]. Copyright 2004, Oxford University Press. Finally, one of several pathways is taken to resolve the intermediate recombination complex.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

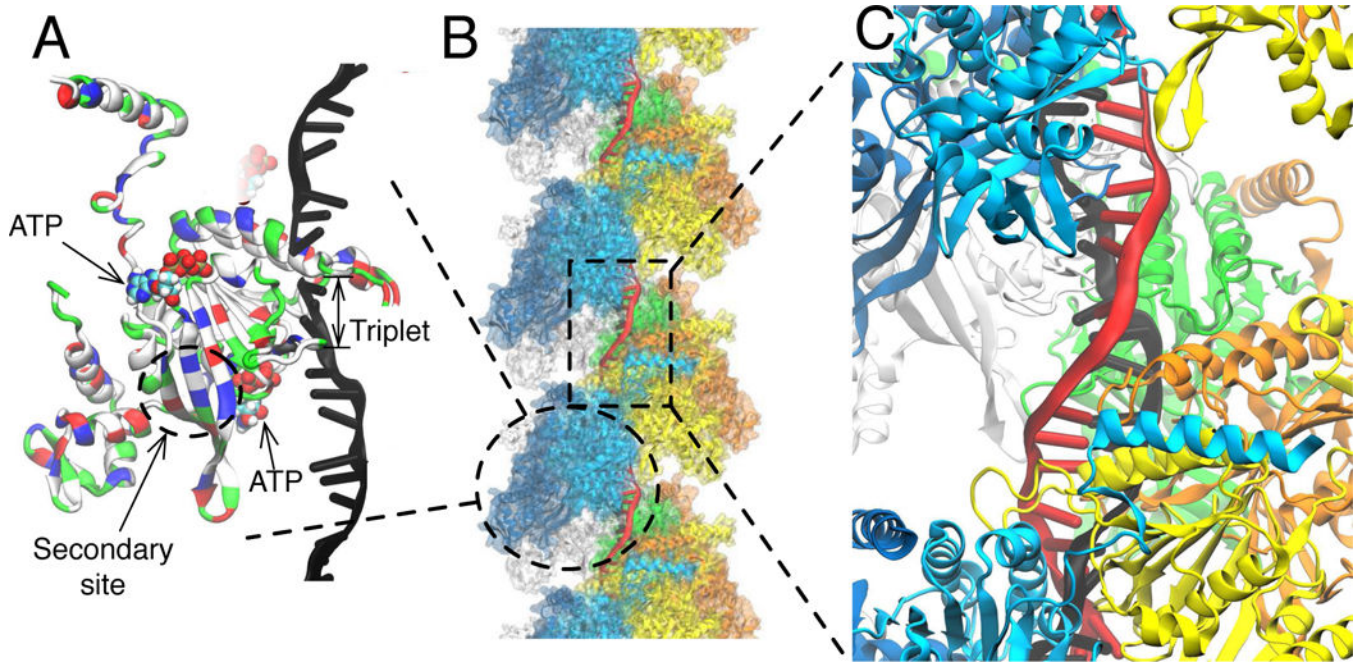


Figure 8:

All-atom structures of RecA. (A) Crystal structure of a RecA monomer shown in a cartoon representation and colored by the residue type: basic (blue); acidic (red), polar (green), and hydrophobic (white). To highlight the DNA binding domain, the reference DNA strand is shown in black. Two ATP molecules are bound to the ATP binding domains. (B) Crystal structure of a RecA filament. Monomers of RecA are shown as molecular surfaces of different colors. The backbones of the reference and incoming DNA strands are shown in gray and red, respectively. (C) Close-up view on the structure of the RecA filament showing the reference DNA strand (gray) bound to the incoming DNA strand (red) [268]. For clarity, only those parts of the proteins that are in immediate contact with the DNA strands are shown using the same color scheme as in panel B.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Modeling and simulation studies in the general area of DNA metabolism organized according to the system type and computational model employed.

Approach	Computational method	DNA replication	DNA repair	DNA recombination
All-atom simulation	**MD simulation**	Pol I [28–32] Pol α [33, 34] DNA helicase [35–37] Pol γ [38] PCNA [39–42] gp45 [39] Pol III [39, 43] TLS [44] SSB [45]	Pol β [46–55] Pol γ [38] Pol η [56] Pol λ [57] RecA [58] Ku [59, 60] DNA glycosylase [61–64] Photolyase [65] PCNA [39–41, 66, 67] gp45 [39] Pol III [39] APE1 [68] Histone ubiquitination [69] 9–1-1 protein [41, 67, 70] SSB [71] AlkB [71] Rad4 [72, 73] Rad32B [74] MuM [75] MutS [76–82] DNA bending/flipping [83–94]	RecA [95, 96] Rad51[97, 98] Homologous DNA [99, 100, 162] Holliday junction [101–104] DNA bending/flipping [105–107]
	Enhanced sampling	Pol I [108] Pol β [109] DNA helicase [110, 111] Pcra [111] ORC/DnaA [112]	Photolyase [113] Pol β [52, 109, 114–117]	Holliday junction [118]
	FEP	Pol β [119–121] T7 DNA Pol [120, 122]	Pol β [46, 55, 123] Pol λ [123]	
	LRA	Pol β [119, 120] T7 DNA Pol [120, 124]	Pol β [46, 125]	
	Docking		Ku [126] Photolyase [127] PCNA [66] SSB [71] AlkB [71] DNAligase [128]	
QM	**QM/MM**	Pol β [129] Pol η [130] RNase H [131–133] hFEN1 [132] Dpo4 [134]	Pol β [54, 115, 129, 135–138] Pol λ [139] Photolyase [140] MutM [141]	
	Ab-initio	DNA-DNA [142–144] DNA polymerase [145] Pol β subunit [146]	DNA-DNA [143, 144, 147–150]	
	EVB	Pol β [120] T7 DNA Pol [120, 122, 151]	Pol β [55, 123] Pol λ [123]	
CG simulation	**MD simulation**	PCNA [40, 152, 153] RPA [154] DnaA [155] SSB [156] Hexameric helicases [157, 158]	PCNA [40, 152] RPA [154] DNA-protein complex [163–165]	Holliday junction [159, 160] Homologous DNA [99] RecA/Rad51 [161]
	Docking	DNA–protein complex [167]		RecA [166]
Stochastic model	**Rate equation**	Chromosome [168, 169]	NHEJ [170, 171]	RecA [172–174] Rad51 [175]
	Stochastic model	Chromosome [176–181] orC [182]	BER enzyme [183–185] DNA glycosylase [62, 63] NHEJ [186, 187]	RecA [188] Rad51 [189] RPA [189]

Adv Theory Simul. Author manuscript; available in PMC 2020 May 01.