Published in final edited form as:

Nanomedicine (Lond). 2008 February ; 3(1): 93–105. doi:10.2217/17435889.3.1.93.

Understanding and Re-engineering Nucleoprotein Machines to Cure Human Disease

William Dynan^{1,†}, Yoshihiko Takeda¹, David Roth², and Gang Bao^{3,†}

¹ Institute of Molecular Medicine and Genetics, Medical College of Georgia, Augusta, GA 30912, USA

² The Kimmel Center for Biology and Medicine of the Skirball Institute of Biomolecular Medicine and Department of Pathology, New York University School of Medicine, New York, NY 10016 USA

³ Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, USA

Summary

The mammalian nucleus is filled with self-organizing, nanometer-scale nucleoprotein machines that carry out DNA replication, RNA biogenesis, and DNA repair. We discuss, as a model, the nonhomologous end-joining (NHEJ) machine, which repairs DNA double-strand breaks. The NHEJ machine consists of six core polypeptides and 10–20 ancillary polypeptides. A full understanding of its design principles will require measuring the behavior of single NHEJ complexes in living cells, using a Nano Toolbox that includes bright, stable, biocompatible fluorophores, efficient protein and nucleic acid tagging strategies, and sensitive, high-resolution imaging methods. Taking inspiration from natural examples, it may be possible to adapt and redesign the NHEJ machine to precisely correct mutations responsible for common human diseases, such as K-ras in lung cancer, or human papillomavirus E6 and E7 genes in cervical and oral cancers.

Keywords

Nonhomologous end joining; DNA-dependent protein kinase; DNA repair; DNA recombination; Ionizing radiation; Gene therapy; Protein labeling; Quantum dots; Super-resolution microscopy

Introduction

The mammalian cell nucleus is a membrane-bounded compartment filled with self-organizing, interconnected, nanometer-scale machines. These machines carry out essential processes of DNA replication, RNA synthesis, pre-mRNA processing, early ribosome biogenesis, RNA transport, and DNA repair [1]. We refer to the general class of machines that are made primarily of proteins and act on nucleic acid substrates as “nucleoprotein machines.” They are complex: synthesis of a typical human mRNA, for example, requires precise interaction of hundreds of protein and RNA components, including initiation, capping, elongation, splicing, polyadenylation and termination factors.

Nucleoprotein machines are quite dynamic and do not have a fixed composition. For example, the nonhomologous end joining (NHEJ) machine that repairs DNA double strand breaks (DSBs) requires a different constellation of ancillary factors depending on the complexity of

[†]Correspondence may be addressed to either WSD (Tel+1 706 721 8756; Fax +1 706 434 6440; E-mail: wdynan@mcc.edu) or GB (Tel 1 404 385 0373; Fax +1 404 894 4243; E-mail: gang.bao@bme.gatech.edu).

the DNA damage [2]. NHEJ also occurs within domains of modified chromatin containing roughly 2 million base pairs of DNA [3] and must be able to join ends of virtually any nucleotide sequence. The uniqueness of individual repair complexes exemplifies why it is essential to study the behavior of single nucleoprotein complexes, rather than population averages, in order to obtain insights into their design principles.

Elucidation of these principles involves understanding: (1) the role of each component in a nanomachine, (2) the pathway by which the machine assembles and disassembles, and (3) the signalling and control mechanism within and between nanomachines. Nucleoprotein machines work with a common set of raw materials (nucleotides and polynucleotides), carry out similar elementary steps (nucleotidyl and phosphoryl group transfer), and often have interchangeable components. TFIIF, for example, participates in both RNA transcription and nucleotide excision repair [4], and Ku and DNA-PKcs participate in both NHEJ and in telomere maintenance [5,6]. If each of these components works by the same mechanism in different processes, then it is likely that study of different machines will reveal common, and generalizable, engineering principles.

A long-term goal is to adapt and redesign nucleoprotein machines to carry out novel functions, including precise modification of the information stored in DNA (or RNA) to provide genetic cures for common human diseases. Consistent with this goal, we focus here on the NHEJ machine that repairs DNA DSBs. This machine has an intrinsic ability to add, delete, and rejoin DNA sequences at the break sites. Adapting this machine to directly manipulate the information encoded in DNA on the nanoscale in a directed fashion would provide a broad and powerful approach to medicine that potentially transcends limitations of present-day pharmaceutical technology.

The NHEJ Complex—a Model Nucleoprotein Machine

Specialized DNA repair machines exist for repair of base damage (base excision repair), bulky nucleotide adducts (nucleotide excision repair), replication errors (mismatch repair), interstrand crosslinks, and DSBs (homologous recombination (HR) and NHEJ) [7]. We focus here on NHEJ as a model system because it is simple, compared to many other nucleoprotein machines, and because it occurs in discrete self-assembling structures that can be visualized by microscopy [3].

NHEJ repairs DSBs, which are among the most potent and threatening types of DNA lesion. Unlike most other forms of DNA damage, DSBs endanger entire chromosomes, not just a local site. Every chromosome is a single, very long molecule of DNA. A DSB divides the one molecule into two, with the potential for loss of one of the fragments or aberrant re-joining to create transposed, dicentric, or ring chromosomes. Despite the existence of efficient repair systems, on average only about 40 DSBs per mammalian cell are needed to lead to cell death. This contrasts with other types of damage, such as base oxidation, alkylation, or photodimerization, where 100,000–800,000 lesions per cell are required for the same level of toxicity (Ward et al., cited in ref [8]). One of the reasons that NHEJ has been so extensively studied is its role in determining the outcome of radiation therapy, a treatment that is received by more than half of all cancer patients. Radiation therapy works by inducing DSBs in tumor cells, and the level of NHEJ enzymes directly determine clinical radioresistance [9].

The catalytic core of the NHEJ machine

Six polypeptides make up the catalytic core of the NHEJ machine (Fig. 1). The two-subunit Ku protein carries out initial recognition of broken DNA ends. It then recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which also binds directly to the DNA ends and regulates access by other proteins. Naturally-occurring DSBs, particularly those

induced by ionizing radiation, frequently contain various types of base and sugar damage that must be processed by additional enzymes. Damaged nucleotides are removed by nucleases, gaps are filled by DNA polymerases [10], and phosphate groups are added and removed by polynucleotide kinase [11] (Table 1). The DNA ligase complex, composed of DNA ligase IV, XRCC4, and (probably) Cernunnos-XLF, then catalyzes phosphodiester bond formation (Table 1).

Mammalian cells that are deficient in any of the core NHEJ components are very sensitive to DSB-inducing agents [12–18]. Complete absence of NHEJ, arising from null mutations in components of the DNA ligase complex, results in embryonic lethality attributable to accumulation of endogenous DNA damage in postmitotic neurons [19]. Even subtle polymorphisms in NHEJ components can affect cancer risk. A hypomorphic DNA-PKcs allele in the mouse, for example, increases the risk of breast cancer by several-fold [20].

V(D)J recombination

One of the attractions of the NHEJ machine as a model is that nature has already adapted the machine for targeted gene mutation. The core NHEJ machine is evolutionarily ancient. It is present in all eukaryotes that have been examined and in many bacterial lineages [21]. NHEJ acquired a new function about 400 million years ago, at the time of the emergence of jawed vertebrates. Introduction of an additional protein component, encoded by the Recombination Activating Genes, Rag1 and Rag2, allowed the NHEJ machine to perform combinatorial joining of germ-line DNA segments to create a wide variety of antigen receptors. This process, which is restricted to lymphocyte lineages, involves Rag-catalyzed cleavage of DNA at recombination signal sequences, followed by joining of Variable, Diversity, and Joining DNA segments (reviewed in [22]). The process is thus known as V(D)J recombination, and it is the only site-specific DNA rearrangement known in vertebrates. V(D)J recombination is essential to make an adaptive immune system capable of recognizing novel pathogens and other foreign molecules. As such, it is a spectacular evolutionary success story. As discussed in a later section, the natural role of NHEJ in V(D)J recombination provides inspiration for the idea that it may be possible to develop engineered biomolecular devices that operate along similar lines, which could function as “gene correction machines” to cure common genetic diseases.

Repair foci

The genetics and biochemistry of the core NHEJ machine and of the RAG proteins are well understood, and high resolution structures are available for some individual components. However, our understanding of how components assemble and interact on a larger scale (roughly 20 to 200 nm) remains quite uncertain. NHEJ occurs *in vivo* in the context of repair foci, which can be detected by immunostaining for a phosphorylated histone variant, γ -H2AX [3,23], and appear initially as point sources at or near the resolution limit of light microscopy [24,25]. Their initial diameter is no more than about ~200 nm. There is strong evidence that immunostained repair foci correspond 1:1 with sites of DNA DSBs [3,26]. In yeast, γ -H2AX phosphorylation has been shown to extend about 50 kb to either side of a defined DSB [27]. In mammalian cells, where an exact measurement has not yet been reported, calculation suggests that the phosphorylation extends on average at least 1 megabase pair (Mbp) [3]. This means that DNA in mammalian foci has a contour length of about 70,000 nm, several hundred times its observed size. Although some chromatin expansion occurs in the damaged region [28], the chromatin evidently remains highly compacted relative to free DNA. Repair foci contain a number of proteins in addition to γ -H2AX, including phosphorylated DNA-PKcs, 53BP1 [29], MDC1/NFBD1 [30], HDAC4 [31,32], and the Mre11•RAD50•NBS1 complex [33] (Table 1). Many of these are involved in cell cycle or other regulatory functions, rather than in catalysis of the core NHEJ reaction. Each is probably present in several dozen copies or more, because they can be detected readily by immunostaining. However, their spatial

arrangement, stoichiometry, and interactions are highly uncertain. Interestingly, although the core components of the NHEJ machine, including Ku and the DNA ligase complex, must be present in the foci (because their activities are essential for NHEJ), they are not detected directly by immunostaining. Presumably they are insufficiently abundant (low signal) or their presence is obscured by abundant free protein present throughout the nucleus (low signal-to-noise ratio).

Nano Toolbox Development

Our current understanding of the NHEJ machine derives from biochemical and genetic studies. Although these approaches continue to be important, achieving the next level of understanding will require complementary methods that can track the assembly of individual protein and DNA molecules, *in vivo*, into repair complexes. This is because each DSB repair complex is to some extent unique: the chemical structure of DNA ends, the DNA sequence flanking those ends, the overall dimensions of the focus, and presumably, the complement of proteins differ in each instance. Current technologies allow tracking of single molecules on cell surfaces, but not, as yet, in a complex natural milieu. Tracking the dynamic behavior of core components of the NHEJ machine inside living cells will therefore require new tools (Box 1).

Fluorescence probes and tagging strategies

To track multiple components of nucleoprotein machines in living cells, each component must be tagged with a spectrally distinct probe. Two general strategies are available, one based on genetic fusion of a fluorescent protein with the target protein, the other based on chemical conjugation with an organic or inorganic fluorophore (Table 2). The first of these strategies, genetic fusion, is widely used because it affords perfect specificity in labeling, controlled 1:1 stoichiometry, and the ability to introduce the fluorophore by the normal pathway of cellular transcription and translation. Currently used fluorescent proteins derive from marine organisms, *Aequoria victoria* and *Discosoma sp.* Reengineering of the natural fluorescent proteins from these organisms has provided a palette of spectrally distinct, monomeric proteins, making it feasible to track up to four target proteins simultaneously [34]. Although photobleaching limits sensitivity for intracellular single-molecule studies, it may be possible to overcome this by improving the efficiency of capture and analysis of emitted photons, and/or the photostability of fluorescent proteins. Another potential drawback of using fluorescent proteins is their steric bulk. *A. victoria* green fluorescent protein, for example, folds into a cylinder 2.4 nm in diameter and 4.2 nm in height [35,36], potentially large enough to interfere with folding of the target protein or its assembly into a larger complex. Examples of fluorescent protein-tagged NHEJ components are shown in Figure 2. Sites of tagging were chosen to minimize potential for interference with function, based on prior mutational and structural data.

A second general labeling strategy, based on chemical conjugation, provides access to fluorophores that are, in some instances, brighter, smaller, or more stable than fluorescent proteins. It has been demonstrated that peptide tags can be labeled with biarsenical dyes (tetracysteine), nickel tris-nitriloacetic acid-conjugated dyes (oligohistidine) (see Table 2 for references), Texas Red (“fluorettes”), or biotin (biotin ligase acceptor peptides). Biotin ligase acceptor peptides are particularly advantageous because biotin ligases recognize cognate acceptor peptides with high specificity, and the use of different biotin ligases with orthogonal recognition specificities permits multiplex labeling [37,38]. Somewhat larger tags, derived from specific receptors and enzymes, have also been developed (Table 2). These afford higher affinity and specificity than peptide tags, at the cost of steric bulk. One interesting scheme uses a 150-residue self-cleaving protein domain (an intein) that removes itself to provide a reactive N-terminal cysteinyl residue, which can be used for chemical ligation. Although the reaction is quite slow, it has the advantage that the bulky tag is actually removed, so that it will not interfere with the function of the labeled protein.

The exogenous fluorophore in these schemes may be either an organic dye or a Quantum Dot (QD). A QD is typically made from a semiconductor nanocrystal with a core-shell structure that confines the motion of conduction band electrons, valence band holes, or excitons in all three spatial directions. Due to this confinement, a QD has a discrete quantized energy spectrum, and very high quantum yield. They are at least 20 times brighter and 100 times more photostable than comparable dyes [39]. They have broad absorption spectra but very sharp emission spectra, which can be continuously tuned by varying the size of the QD over a range of 2–10 nm. Thus, a strategy based on QDs can be used to track 10 or more labeled proteins simultaneously using a single excitation wavelength. Unfortunately, steric bulk of currently available QDs is a significant drawback. With a diameter of 15–25 nm, they occupy a 100–400 fold greater volume than a folded GFP moiety. Most of this bulk is attributable to biocompatible coatings, such as dihydrolipoic acid, hydrophilic organic dendron ligands, or streptavidin-conjugated amphiphilic polymer, that enable the use of QDs in an aqueous environment [40]. It may be possible to develop thinner biocompatible coatings, such as histidine-tagged peptides (10–30 residues) [41]. Although QDs can be delivered into living cells using, for example, cell penetrating peptides or streptolysin O (SLO) [42,43], they cannot be washed out once delivered into cells. Therefore, it is critical to optimize the concentration of QD bioconjugates that are delivered into living cells to increase the signal-to-noise ratio. An alternative is to perform the QD conjugation *ex vivo* and load the entire QD-tagged proteins into cells by microinjection or other methods.

Controlled induction of single DSBs

One of the most widely used methods for inducing DSBs involves low-dose exposure to sparsely ionizing radiation such as γ -rays, which transfer energy to solvent (water) along nanometer-scale tracks, creating a trail of hydroxyl radicals and other reactive oxygen species. When a radiation track intersects a DNA helix, clustered damage occurs, which leads directly or indirectly to DSB formation. The amount of radiation required to induce a given number of DSBs is accurately known. For example, a gamma-ray dose of 1 Gray (1 Gray = 1 J/kg absorbed energy) will produce approximately 30 DSBs per human genome equivalent [44,45]. This dose can be delivered in ≤ 1 minute using a ^{137}Cs laboratory irradiator. Induction of DSBs is stochastic, such that many types of local sequence environments and chemical structures are represented in the population of DNA ends.

A more selective, sequence-specific approach for DSB induction is based on rare-cutting nucleases. In yeast, the well-characterized HO endonuclease introduces a DSB at a single unique site [46]. In mammals, the task is harder because of the much larger genome and the absence of natural enzyme systems for efficient formation of DSBs. Transient transfection has been widely used to express the cDNA for the homing nuclease, I-SceI, in human cells, resulting in the induction of DSBs on a time scale of hours to days. I-SceI has an asymmetric recognition site of ~40 nt (although imprecise specificity causes it to behave as if the site size is ~12 nt, occurring by chance 1 in every 16 million bp). Expression of transfected I-SceI is useful if the readout involves stable genome modification, but the time scale of expression is too slow for dynamic studies of repair complex assembly. Very recently, the Misteli group has overcome this problem by fusing I-SceI to a glucocorticoid receptor binding domain to create an enzyme that can be rapidly activated in the presence of ligand [47]. Striking induction of DSBs occurs over a time scale of just a few minutes (Fig. 3). This system holds great promise for single-molecule studies of DSB repair complexes *in vivo*.

Imaging methods with high resolution, sensitivity, and signal-to-noise ratio

Three factors limit the ability to visualize single-molecule processes *in vivo* by optical microscopy: (1) sensitivity, (2) photobleaching and phototoxicity (mediated by endogenous chromophores), and (3) optical resolution. Improvements in sensitivity will come primarily

from improvements in fluorescence probes and tagging strategies discussed in preceding sections. QDs, for example, are bright enough that a single QD can easily be distinguished against background. Photobleaching and phototoxicity can be minimized by using brighter and more stable probes and by developing instrumentation that makes efficient use of emitted photons. With ordinary confocal microscopy, the entire specimen is illuminated, but information is collected only from the focal slice. With multiphoton confocal microscope, excitation occurs through a nonlinear mechanism only at the convergence of two laser beams, minimizing wasteful illumination of other parts of the specimen. With deconvolution microscopy, which is another method for obtaining z-axis resolution, the entire specimen is illuminated, all of the emitted photons are collected (both in-focus and out-of-focus), and computational methods are used to deconvolve emissions from individual optical slices along the z-axis. Because of efficient photon use, only very short illumination times are required.

Optical resolution is a more challenging issue. Conventional and multiphoton confocal microscopy, as well as deconvolution microscopy, are limited by the inability to resolve objects separated by less than about half of the wavelength of light used to observe the object (the “Rayleigh diffraction limit”). This is a significant problem in the case of repair foci having dimensions of 200 nm or less. Several approaches have been taken to develop so-called “super-resolution” optical microscopy (reviewed in [48]). Box 2 lists several that are applicable in living cells. Structured Illumination Microscopy (SIM) and Stimulated Emission-Depletion (STED) involve manipulation of sample illumination, in one case combining interference patterns from multiple exposures, and in the other using a combination of two shaped laser pulses. Photoactivated Localization Microscopy (PALM) and Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM) exploit properties of single-molecule behavior. None of the methods has yet been applied to living cells, where phototoxicity and sample motion may limit their use. However, rapid fixation methods, in combination with bright stable fluorophores, orthogonal tagging methods, and super-resolution microscopy, hold great promise for investigating the assembly pathway for individual nucleoprotein machines.

Understanding the Engineering Design of the NHEJ Machine

How will the Nano Toolbox extend the knowledge gained from conventional biochemical, genetic, and cell biological studies? The design of nucleoprotein machines in living cells has been optimized by nature over more than a billion years of evolution. In comparison to devices designed by human engineers, nanomachines realize their functions using relatively few components and with astonishing precision, efficiency, and robustness. For example, DNA repair machines can recognize defective nucleotide residues in a background of more than 10^9 normal residues and in many instances correct them with accuracies on the order of 99.99% (i.e., induced mutation rates on the order of 10^{-4}). In order to uncover the design principles of a biological nanomachine, we need to understand its structure-function relationships (design blueprints) and generate a quantitative description of its behavior. Single molecule studies using tagged components will provide the opportunity to address:

- How fast the machine runs (kinetics). For a single DSB at a defined site, it will be possible to use tagged Ku, DNA-PKcs, and DNA ligase complex to determine the order of assembly, the time interval between recruitment of one component and the next, and importantly, whether these events occur in the same order and with the same timing at each break.
- How accurately the machine works (accuracy/sensitivity). For breaks introduced at a defined site, it will be possible to quantify the accuracy of repair by measuring the susceptibility to re-cleavage, or by sequencing individual products to determine the frequency of insertions and deletions.

- How quickly the machine changes its form/shape for different tasks (robustness). Many repair proteins are multifunctional. Assembly of DSB repair complex may be controlled, in part, by the rate of release of NHEJ proteins from other sites or storage locations which can be measured using the same approaches discussed in preceding sections.
- How well the machine is controlled (feedback). The signaling protein, MDC1, provides an example of the type of control circuitry that affects DSB response in vivo [49]. The presence of broken DNA ends promotes initial activation of the ATM kinase, which modifies the chromatin in the vicinity of the DSB. This promotes binding of the signaling protein, MDC1, which in turn activates more ATM, amplifying the original signal. At present, we have only a qualitative understanding of this and other aspects of the regulatory circuitry. Quantitative, time-resolved data describing growth and decay of repair foci will allow for a more accurate description of the control circuitry.

Although it is difficult to predict the types of design principles that may emerge from these studies, we present several hypotheses:

1. Useful engineered nanomachines will have a mechanism to physically and topologically isolate their DNA substrate. The Rag proteins accomplish this by forming a DNA loop containing the intervening sequence that is to be removed from the genome when V, D, or J segments undergo joining. Other tightly regulated systems, including type II restriction endonucleases, bacterial repressors, and mammalian transcription factors, may follow a similar strategy.
2. Different nanomachines will have interchangeable parts, i.e., small protein assemblies that execute the same preset series of actions. TFIIH performs the same task (making a DNA bubble) in nucleotide excision repair and transcription, and PSF•p54 promotes pairing of distant sequences in many contexts.
3. The repeating polymeric structure of DNA and RNA can be used to form natural biological amplifiers. A signal initiating at a single site propagates linearly via chromatin modification, providing a long dock for signaling proteins that arrive, undergo modification, and depart to propagate the signal three-dimensionally. The 70,000 nm of DNA contour in mammalian repair foci might accommodate thousands of signaling proteins, with a corresponding gain in potential signal strength.

Pathway to Medicine

The term “Pathway to Medicine” has been coined to describe efforts to translate basic work in nanomedicine for clinical application. We describe here a conceptual approach toward creating “gene correction machines” by re-engineering natural repair and recombination processes. We conceptualize this pathway in terms of three elements: *device*, *delivery*, and specific *disease* targets.

Device

For the *device*, we draw inspiration from a natural example. The Rag proteins incise the genome at two specific points, hold the broken ends together, and hand off the four resulting ends to the NHEJ machine [22,50]. This V(D)J recombination event lies at the heart of the adaptive immune response – the process by which exposure to an infectious agent, or a vaccine, results in lifelong immunity. By analogy, we envision the ability to create an engineered nanodevice, composed of protein, nucleic acids, lipids, artificial materials, or a composite of all of these, which would enter cells, interface with the DSB repair machine, and achieve a targeted modification. There are a number of human diseases that are associated with a dominant,

somatically acquired single-gene mutation. Excision of the mutant allele, in a reaction analogous to cutting and pasting in V(D)J recombination, would provide a genetic cure for the diseases. Through this approach, it might be possible to correct errors in the somatic genome through a procedure that is as safe, inexpensive, and acceptable as vaccination is today.

Is it feasible to develop a device that, like the Rag proteins, will bind, incise, tether, and ultimately remove a DNA segment within a pre-selected gene of interest? One approach is to re-engineer the Rag proteins themselves to recognize a different, pre-selected sequence. This re-engineered protein would then coordinate the remaining steps of the recombination reaction. An alternative is to start with a DNA binding protein of predetermined specificity and engraft functionality required to incise the DNA ends, tether them, and recruit the NHEJ machinery. Zinc finger proteins can be designed to recognize pre-selected sites in the genome, for example, and readily coupled to protein domains that supply additional functionalities [51]. The ability to implement these approaches is currently limited by our incomplete knowledge of the Rag complex; they will be more feasible when a high-resolution structure is available and the DNA and protein recognition interfaces are better defined.

We have recently become interested in another approach that may be characterized as a hybrid of the V(D)J/NHEJ and homologous recombination pathways. Mammalian Rag proteins are thought to be coupled exclusively to the NHEJ pathway. However, introduction of a point mutation into one of the Rag polypeptides partially disables the active site, causing the protein to make single-strand rather than double-strand breaks. Surprisingly, this simple mutation causes Rag to promote entry of nicked DNA intermediates into the homologous recombination (HR) pathway instead of NHEJ [52]. Thus, Rag initiates gene conversion rather than a sequence deletion. Fig. 4 shows how this approach might be implemented: a DNA minicircle bears a fragment of a wild-type gene and recombination signal sequence (a recognition site for Rag complex). Upon entry into the cell and uncoating, the Rag protein introduces a site-specific nick. It is presumed that a single strand of DNA is extruded from the nicked minicircle, invades the mutant genomic locus, and promotes the recombinational repair of the mutant gene. The strategy circumvents the need to introduce free double-strand ends in either the device or the genome, thus precluding any possibility of random insertion or gene rearrangement. Also, the device as depicted in Fig. 4 involves only proteins and nucleic acids that are normally present in the human body, making it unlikely that an immune response would compromise its function. Preliminary proof-in-principle of the feasibility of such a device has been demonstrated in cell culture using a model fluorescent protein gene [52]. A challenge now is to demonstrate that the approach can be delivered *in vivo* to correct genes involved in human diseases.

Delivery

To be useful, an engineered nanodevice must (1) reach its target tissue, (2) enter the target cell, (3) evade sequestration in intracellular vesicles, and (4) enter the appropriate cellular compartment. Given these requirements, oncology applications are particularly attractive. Leaky tumor vasculature allows nanoparticles to escape the bloodstream; indeed there is already an FDA-approved nanoparticle therapy based on a liposome-encapsulated form of the DNA-damaging drug, doxorubicin [53]. Folate receptor (FR)-mediated endocytosis has shown promise for promoting entry of therapeutic macromolecules into cells [54]. Folate is essential for thymidylate synthesis, and, therefore, for DNA replication. Cancer cells often express high levels of FR [55]. We hypothesize that folate-derivatized nanodevices should be delivered effectively, and with some selectivity, to cancer cells. Targeted delivery may not be essential (a device that is designed to correct a mutant oncogene, would be intrinsically selective for tumor cells, because only tumor cells bear the mutation) but would contribute to the efficiency of the treatment.

FR-mediated endocytosis delivers molecules into intracellular vesicles, where they may become trapped and thus unavailable. Viruses, which must overcome the same problem, have evolved natural strategies for vesicle disruption. Influenza virus, for example, has a capsid protein, hemagglutinin, that undergoes a conformational change at the characteristically low pH encountered in intracellular vesicles. A short hemagglutinin-derived peptide is sufficient to mimic this function and promote cytoplasmic release of foreign macromolecules [56]. Targeting to the appropriate intracellular compartment may be the most easily solved part of the deliver problem. Nuclear localization peptides promote nuclear uptake of particles as big as QDs and should readily promote entry of a smaller nucleoprotein machine [57].

In summary, technologies exist to breach each of the multiple barriers that separate an exogenous nucleoprotein from its intended target. Successful delivery of a nanodevice will not require discovery of new scientific principles; rather delivery is an engineering problem that will require combination and refinement of existing technologies.

Disease

The best prospects for translation of the gene correction machine concept involve diseases (1) that are common, (2) for which the consequences of disease are severe (i.e., life-threatening), and (3) that are associated with a single, well-defined mutant allele. We will mention here two examples from the oncology field: Mutant K-ras-associated cancers of the lung, pancreas, and colon, and human papillomavirus-associated cancers of the uterine cervix and of the head and neck.

K-ras is one of the most frequently mutated genes in human cancers. A member of the G-protein family, K-ras is involved in transducing growth-promoting signals from the cell surface. Point mutations of K-ras have been found in 80–100% of pancreatic, 40–60% of colon, and 25–50% of nonsmall cell lung cancers. All three of these cancers are common (lung cancer is the most common cause of death worldwide). Except for early-stage colorectal cancer, which can be cured surgically, the odds of long-term survival are unfavorable. The mutant allele is well defined, as K-ras mutations occur almost exclusively in three hot spots (codons 12, 13 and 61). In pancreatic cancer, K-ras mutations occur very early in malignant progression, such that a strategy of targeting K-ras mutations can lead to a better outcome. In lung cancer, K-ras mutation is associated both with radioresistance ([58]) and with primary resistance to new, targeted chemotherapy agents, gefitinib (Iressa) and erlotinib (Tarceva) [59], underscoring its clinical significance.

Human papillomavirus (HPV) is the causative agent of cervical cancer, the second most common cause of cancer death among women worldwide [60]. HPV is only one of several factors involved in the etiology of head and neck cancers, which account for about 125,000 deaths annually [61]. However, there is strong evidence that infection increases cancer risk, and viral DNA is present in about 30% of cases in Western countries [62–64]. For both cervical and oral squamous cell cancers, the prognosis for advanced disease is unfavorable and existing treatments fall short of a cure. Expression of HPV E6 and E7 oncogene is essential for growth of transformed cells, making these genes excellent targets for introduction of a mutation that abrogates expression.

Future perspectives

We present the NHEJ machine here as a model because it is relatively simple and easy to visualize using optical microscopy. However, the cell nucleus is filled with other interconnected, nanometer-scale machines that carry out essential processes of information storage and decoding. The Nano Toolbox can readily be adapted to study these other nucleoprotein machines. The mismatch repair machine, for example, involves a different set

of gene products than NHEJ, but shares with NHEJ the ability to introduce precise alterations in genomic DNA sequences. Like NHEJ, mismatch repair might be re-engineered to address common human genetic disorders. A greater challenge is posed by the RNA transcription and splicing machines, which have hundreds of components, yet could also be approached using the Nano Toolbox.

We have not discussed here the topic of mathematical modeling of nucleoprotein machine dynamics and control. We anticipate, however, that single-molecule studies of the NHEJ machine can be used as input parameters for robust, predictive models. Such models may be useful for drug development, predicting, for example, which components in a nucleoprotein machine can be targeted to best disrupt or redirect its activity.

Finally, we have limited our discussion of engineered biomolecular devices to those fabricated from natural materials, such as proteins, nucleic acids, and lipids. Clearly, however, incorporation of non-natural materials, such as metals, semiconductors, or novel organic compounds, would lend additional versatility and foster the ability to direct novel processes inside living cells and organisms.

Executive summary

- The term “nucleoprotein machines” refers to a class of nanomachines that are made primarily of protein and that act on nucleic acid substrates.
- Essential processes carried out by nucleoprotein machines include DNA and RNA synthesis, RNA processing and transport, and DNA repair.
- Several lines of evidence suggest that different nucleoprotein machines are built according to common design principles.
- The nonhomologous end joining (NHEJ) machine, which repairs DNA double-strand breaks (DSBs) induced by radiation therapy and other causes, is a tractable and clinically significant model nucleoprotein machine.
- Tracking the assembly of NHEJ complexes deep within living cells will require a Nano Toolbox consisting of new fluorescence probes, protein tagging strategies, methods of controlled induction of double-strand breaks, and sensitive, high-resolution imaging.
- The natural process of V(D)J recombination, which uses the NHEJ machine to create antigen receptors in the immune system, provides inspiration for development of an analogous biomolecular device—a gene correction machine.

Box 1

Four challenges to overcome for single-molecule studies in living cells

- Develop high-contrast, stable, fluorescent probes that can be detected on a single-molecule basis.
- Develop orthogonal tagging strategies so that probes can be stoichiometrically conjugated to multiple nonhomologous end joining components. Tags must not interfere with nanomachine assembly.
- Develop efficient methods for controlled induction of single DSBs, preferably at predetermined genomic sites.
- Develop imaging methods with sufficiently high resolution, sensitivity, and signal-to-noise ratio to detect single molecules deep inside living cells.

- Develop computational models to describe the kinetics of nanomachine assembly and disassembly.

Box 2

Super-resolution microscopy

Optical resolution is limited by the width of the point-spread function for emitted light, and in practice is about 200 nm in the XY plane. Strategies to resolve smaller features include:

- SIM (Structured Illumination Microscopy). Uses spatially structured light, which makes high-resolution information visible as moiré fringes [65–67]. Practical resolution 100 nm, with <50 nm feasible.
- STED (Stimulated Emission-Depletion). Relies on two beams: one excites all fluorophores in the illuminated region, the other, at a different wavelength containing a node of zero intensity at its center, suppresses fluorescence except in a 25–50 nm region [68] Practical resolution about 60 nm, with better resolution possible.
- PALM (Photoactivated Localization Microscopy) [69]. Relies on serial photoactivation and bleaching of many sparse subsets of photoactivatable fluorescent proteins. Currently limited to fixed sections and a small palette of photoactivatable fluorescent proteins. Practical resolution 10 nm.
- STORM (Sub-diffraction-limit Imaging by Stochastic Optical Reconstruction Microscopy). [70]. Relies on switching individual dye molecules on and off. Resolution of 20 nm demonstrated for biochemical preparations, not tested with cells.

Acknowledgments

We thank Mr. Kyungsoo Ha and Dr. Katsuya Miyake for help in making and characterizing the fluorescent constructs in Fig. 2. We thank Dr. Tom Misteli for permission to show data in Figure 3. We thank the other investigators in the Nanomedicine Development Center for Nucleoprotein Machines for their contribution to the ideas expressed in this review, including Roland Eils, Hernan Flores-Rozas, Grant Jensen, Shuming Nie, David Spector, and Alice Ting. We thank Dr. Rhea-Beth Markowitz for editorial assistance.

This work was funded by the NIH Roadmap Initiative in Nanomedicine through a Nanomedicine Development Center award, 1PN2EY018244 (GB, WSD, DR). WSD received support from a US Department of Health and Human Services research grant, 5R01CA098239, and is an Eminent Scholar of the Georgia Research Alliance.

Bibliography

1. Spector DL. The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem* 2003;72:573–608. [PubMed: 14527325]
2. Jeggo PA, Lobrich M. Artemis links ATM to double strand break rejoining. *Cell Cycle* 2005;4 (3): 359–362. [PubMed: 15684609]
3. Rogakou EP, Boon C, Redon C, Bonner WM. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* 1999;146 (5):905–916. [PubMed: 10477747]
4. Drapkin R, Reardon JT, Ansari A, et al. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 1994;368 (6473):769–772. [PubMed: 8152490]
5. Bailey SM, Meyne J, Chen DJ, et al. DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. *Proc Natl Acad Sci U S A* 1999;96 (26):14899–14904. [PubMed: 10611310]

6. Hsu HL, Gilley D, Blackburn EH, Chen DJ. Ku is associated with the telomere in mammals. *Proc Natl Acad Sci U S A* 1999;96 (22):12454–12458. [PubMed: 10535943]
7. Nickoloff, JA.; Hoekstra, MF. DNA Damage and Repair. Humana Press; Totowa, N.J: 1998.
8. Background Information and Scientific Principles. Health Effects of Exposure to Low Levels of Ionizing Radiation: BEIR V. National Academies Press; Washington: 1990. p. 9-64.
9. Sirzen F, Nilsson A, Zhivotovsky B, Lewensohn R. DNA-dependent protein kinase content and activity in lung carcinoma cell lines: correlation with intrinsic radiosensitivity. *Eur J Cancer* 1999;35 (1):111–116. [PubMed: 10211098]
10. Mahajan KN, Nick McElhinny SA, Mitchell BS, Ramsden DA. Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* 2002;22 (14):5194–5202. [PubMed: 12077346]
11. Chappell C, Hanakahi LA, Karimi-Busheri F, Weinfeld M, West SC. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J* 2002;21 (11):2827–2832. [PubMed: 12032095]
12. Taccioli GE, Gottlieb TM, Blunt T, et al. Ku80: product of the XRCC5 gene and its role in DNA repair and recombination. *Science* 1994;265:1442–1445. [PubMed: 8073286]
13. Li Z, Otevrel T, Gao Y, et al. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* 1995;83 (7):1079–1089. [PubMed: 8548796]
14. Gu Y, Jin S, Gao Y, Weaver DT, Alt FW. Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proc Natl Acad Sci U S A* 1997;94 (15):8076–8081. [PubMed: 9223317]
15. Gao Y, Chaudhuri J, Zhu C, Davidson L, Weaver DT, Alt FW. A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination. *Immunity* 1998;9 (3):367–376. [PubMed: 9768756]
16. Gu Y, Seidl KJ, Rathbun GA, et al. Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity* 1997;7 (5):653–665. [PubMed: 9390689]
17. Taccioli GE, Amatucci AG, Beamish HJ, et al. Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity* 1998;9 (3):355–366. [PubMed: 9768755]
18. Gao Y, Sun Y, Frank KM, et al. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* 1998;95 (7):891–902. [PubMed: 9875844]
19. Barnes DE, Stamp G, Rosewell I, Denzel A, Lindahl T. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr Biol* 1998;8:1395–1398. [PubMed: 9889105]
20. Okayasu R, Suetomi K, Yu Y, Silver A, Bedford JS, Cox R, Ullrich RL. A deficiency in DNA repair and DNA-PKcs expression in the radiosensitive BALB/c mouse. *Cancer Res* 2000;60:4342–4345. [PubMed: 10969773]
21. Aravind L, Koonin EV. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 2001;11 (8):1365–1374. [PubMed: 11483577]
22. Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 2002;71:101–132. [PubMed: 12045092]
23. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 1998;273 (10):5858–5868. [PubMed: 9488723]
24. Bewersdorf J, Bennett BT, Knight KL. H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy. *Proc Natl Acad Sci U S A* 2006;103 (48):18137–18142. [PubMed: 17110439]
25. Costes SV, Boissiere A, Ravani S, Romano R, Parvin B, Barcellos-Hoff MH. Imaging features that discriminate between foci induced by high- and low-LET radiation in human fibroblasts. *Radiat Res* 2006;165 (5):505–515. [PubMed: 16669704]
26. Rothkamm K, Löbrich M. Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses. *Proc Natl Acad Sci U S A* 2003;100:5057–5062. [PubMed: 12679524]

27. Shroff R, Arbel-Eden A, Pilch D, et al. Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Curr Biol* 2004;14 (19):1703–1711. [PubMed: 15458641]
28. Kruhlak MJ, Celeste A, Dellaire G, et al. Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *J Cell Biol* 2006;172 (6):823–834. [PubMed: 16520385]
29. Schultz LB, Chehab NH, Malikzay A, Halazonetis TD. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J Cell Biol* 2000;151 (7):1381–1390. [PubMed: 11134068]
30. Stewart GS, Wang B, Bignell CR, Taylor AM, Elledge SJ. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* 2003;421 (6926):961–966. [PubMed: 12607005]
31. Xu X, Stern DF. NFB1/MDC1 regulates ionizing radiation-induced focus formation by DNA checkpoint signaling and repair factors. *Faseb J* 2003;17 (13):1842–1848. [PubMed: 14519663]
32. Kao GD, McKenna WG, Guenther MG, Muschel RJ, Lazar MA, Yen TJ. Histone deacetylase 4 interacts with 53BP1 to mediate the DNA damage response. *J Cell Biol* 2003;160 (7):1017–1027. [PubMed: 12668657]
33. Mirzoeva OK, Petrini JH. DNA damage-dependent nuclear dynamics of the Mre11 complex. *Mol Cell Biol* 2001;21 (1):281–288. [PubMed: 11113202]
34. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat Methods* 2005;2 (12):905–909. [PubMed: 16299475]
35. Yang F, Moss LG, Phillips GN Jr. The molecular structure of green fluorescent protein. *Nat Biotechnol* 1996;14 (10):1246–1251. [PubMed: 9631087]
36. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 1996;273 (5280):1392–1395. [PubMed: 8703075]
37. Howarth M, Takao K, Hayashi Y, Ting AY. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* 2005;102:7583–7588. [PubMed: 15897449]
38. Chen I, Choi YA, Ting A. Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins. *J Am Chem Soc* 2007;129 (129):6619–6625. [PubMed: 17472384]
39. Chan WC, Nie S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* 1998;281 (5385):2016–2018. [PubMed: 9748158]
40. Wang YA, Li JJ, Chen H, Peng X. Stabilization of inorganic nanocrystals by organic dendrons. *J Am Chem Soc* 2002;124:2293–2298. [PubMed: 11878983]
41. Xing YCQ, Shen C, Kong KY, Zhau HE, Chung LW, Petros JA, O'Regan RM, Yezhelyev MV, Simons JW, Wang MD, Nie S. Bioconjugated quantum dots for multiplexed and quantitative immunohistochemistry. *Nat Protoc* 2007;2 (5):1152–1165. [PubMed: 17546006]
42. Delahanty JB, Medintz IL, Pons T, Brunel FM, Dawson PE, Mattoussi H. Self-assembled quantum dot-peptide bioconjugates for selective intracellular delivery. *Bioconjug Chem* 2006;17:920–927. [PubMed: 16848398]
43. Duan H, Nie S. Cell-penetrating quantum dots based on multivalent and endosome-disrupting surface coatings. *J Am Chem Soc* 2007;129 (11):3333–3338. [PubMed: 17319667]
44. Iliakis GE, Cicilioni O, Metzger L. Measurement of DNA double-strand breaks in CHO cells at various stages of the cell cycle using pulsed field gel electrophoresis: calibration by means of 125I decay. *Int J Radiat Biol* 1991;59 (2):343–357. [PubMed: 1671686]
45. Ruiz de Almodovar JM, Steel GG, Whitaker SJ, McMillan TJ. A comparison of methods for calculating DNA double-strand break induction frequency in mammalian cells by pulsed-field gel electrophoresis. *Int J Radiat Biol* 1994;65 (6):641–649. [PubMed: 7912713]
46. Haber JE. Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu Rev Genet* 1998;32:561–599. [PubMed: 9928492]
47. Soutoglou E, Dorn JF, Sengupta K, et al. Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* 2007;9 (6):675–682. [PubMed: 17486118]
48. Hell SW. Far-field optical nanoscopy. *Science* 2007;316 (5828):1153–1158. [PubMed: 17525330]
49. Lou Z, Minter-Dykhouse K, Franco S, et al. MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol Cell* 2006;21 (2):187–200. [PubMed: 16427009]

50. Corneo B, Wendland R, Deriano L, Cui X, Kelin A, Roth DB. Rag mutations reveal robust alternative end joining. *Nature*. 2007 In press.
51. Durai S, Mani M, Kandavelou K, Wu J, Porteus MH, Chandrasegaran S. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic acids research* 2005;33 (18):5978–5990. [PubMed: 16251401]
52. Lee GS, Neiditch MB, Salus SS, Roth DB. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* 2004;117:171–184. [PubMed: 15084256]
53. Doxil® (doxorubicin HCl liposome injection), Ortho Biotech Products L.P., Raritan, NJ, Vol. [package insert] (2001).
54. Pan XaL RJ. Tumour-selective drug delivery via folate receptor-targeted liposomes. *Expert Opin Drug Deliv* 2004;1:7–17. [PubMed: 16296717]
55. Parker N, Turk MJ, Westrick E, Lewis JD, Low PS, Leamon CP. Folate receptor expression in carcinomas and normal tissues determined by a quantitative radioligand binding assay. *Anal Biochem* 2005;338 (2):284–293. [PubMed: 15745749]
56. Wadia JS, Stan RV, Dowdy SF. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat Med* 2004;10 (3):310–315. [PubMed: 14770178]
57. Hoshino A, Fujioka K, Oku T, et al. Quantum dots targeted to the assigned organelle in living cells. *Microbiol Immunol* 2004;48 (12):985–994. [PubMed: 15611617]
58. McKenna WGAM RJ. Targeting tumor cells by enhancing radiation sensitivity. *Genes Chromosomes Cancer* 2003;38:330–338. [PubMed: 14566853]
59. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, Zakowski MF, Heelan RT, Kris MG, Varmus HE. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med* 2005;2:e17. [PubMed: 15696205]
60. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of pathology* 1999;189 (1):12–19. [PubMed: 10451482]
61. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA: a cancer journal for clinicians* 2005;55 (2):74–108. [PubMed: 15761078]
62. Silva CE, Silva ID, Cerri A, Weckx LL. Prevalence of human papillomavirus in squamous cell carcinoma of the tongue. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2007
63. Furniss CS, McClean MD, Smith JF, et al. Human papillomavirus 16 and head and neck squamous cell carcinoma. *International journal of cancer* 2007;120 (11):2386–2392.
64. D'Souza G, Kreimer AR, Viscidi R, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *The New England journal of medicine* 2007;356 (19):1944–1956. [PubMed: 17494927]
65. Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy* 2000;198 (Pt 2):82–87. [PubMed: 10810003]
66. Gustafsson MG. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc Natl Acad Sci U S A* 2005;102 (37):13081–13086. [PubMed: 16141335]
67. Heintzmann R, Jovin TM, Cremer C. Saturated patterned excitation microscopy--a concept for optical resolution improvement. *Journal of the Optical Society of America* 2002;19 (8):1599–1609. [PubMed: 12152701]
68. Hell SW. Toward fluorescence nanoscopy. *Nat Biotechnol* 2003;21:1347–1355. [PubMed: 14595362]
69. Qu X, Wu D, Mets L, Scherer NF. Nanometer-localized multiple single-molecule fluorescence microscopy. *Proc Natl Acad Sci U S A* 2004;101 (31):11298–11303. [PubMed: 15277661]
70. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006;3 (10):793–795. [PubMed: 16896339]
71. Chen IaT AY. Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotech* 2005;16:35–40. [PubMed: 15722013]

72. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 2004;22 (12):1567–1572. [PubMed: 15558047]
73. Matz MV, Fradkov AF, Labas YA, et al. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* 1999;17 (10):969–973. [PubMed: 10504696]
74. Wang L, Tsien RY. Evolving proteins in mammalian cells using somatic hypermutation. *Nat Protoc* 2006;1 (3):1346–1350. [PubMed: 17406421]
75. Lata S, Gavuta M, Tampé R, Piehler J. Specific and stable fluorescence labeling of histidine-tagged proteins for dissecting multi-protein complex formation. *J Am Chem Soc* 2006;128:2365–2372. [PubMed: 16478192]
76. Krishnan B, Szymanska A, Gierasch LM. Site-specific fluorescent labeling of poly-histidine sequences using a metal-chelating cysteine. *Chem Biol Drug Des* 2007;69 (1):31–40. [PubMed: 17313455]
77. Rozinov MN, Nolan GP. Evolution of peptides that modulate the spectral qualities of bound, small-molecule fluorophores. *Chem Biol* 1998;5:713–728. [PubMed: 9862799]
78. Gaietta G, Deerinck TJ, Adams SR, Bouwer J, Tour O, Laird DW, Sosinsky GE, Tsien RY, Ellisman MH. Multicolor and electron microscopic imaging of connexin trafficking. *Science* 2002;296:503–507. [PubMed: 11964472]
79. Martin BR, Giepmans BN, Adams SR, Tsien RY. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* 2005;23:1308–1314. [PubMed: 16155565]
80. Miller LW, Cai Y, Sheetz MP, Cornish VW. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat Methods* 2005;2:255–257. [PubMed: 15782216]
81. Marks KM, Braun PD, Nolan GP. A general approach for chemical labeling and rapid, spatially controlled protein inactivation. *Proc Natl Acad Sci U S A* 2004;101:9982–9987. [PubMed: 15218100]
82. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 2002;21:86–89. [PubMed: 12469133]
83. Keppler A, Pick H, Arrivoli C, Vogel H, Johnsson K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* 2004;101 (27):9955–9959. [PubMed: 15226507]
84. George N, Pick H, Vogel H, Johnsson N, Johnsson K. Specific Labeling of Cell Surface Proteins with Chemically Diverse Compounds. *J Am Chem Soc* 2004;126:8896–8897. [PubMed: 15264811]
85. Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol Cell* 2006;22 (4):511–519. [PubMed: 16713581]

Figure 1. NHEJ complex formation in mammalian cells

The Ku70/Ku80 heterodimer forms a ring that preferentially binds to DNA ends and recruits DNA-PKcs. Processing factors (nucleases, DNA polymerases, and polynucleotide kinase) and the DNA ligase complex (DNA Ligase IV, XRCC4, and XLF) then complete the reaction. The geometry of the Ku-DNA-PKcs complex is based on a cryo-electron microscopy study of reconstituted repair complexes [85], but the actual number, arrangement, and order of assembly of components in the *in vivo* repair complex is unknown.

A. Palette of spectrally distinct fluorescent proteins

		Excitation	Emission
ECFP (cyan)		434	477
EYFP (yellow/green)		514	527
mCherry		587	610
mPlum		590	649

B. Examples of tagged NHEJ proteins

Figure 2. Fluorescent protein tagging strategies

(A) Spectral characteristics of four proteins that can be used for fluorescence visualization of components of the nucleoprotein machine. Data from ref [34]. (B) Examples of fluorescently tagged NHEJ proteins. Proteins were expressed transiently in HeLa cells under control of a strong constitutive promoter. Images were collected by multiphoton scanning confocal microscopy and have been merged with a differential interference contrast image of the same field. Details of constructs will be published elsewhere.

Figure 3. Inducible formation of DSBs

An I-SceI site, flanked on one side by tandem lac operators and on the other by tandem tet operators, was introduced into mouse fibroblasts. (A) Cells transfected with Cyan Fluorescent Protein-lac repressor and Red Fluorescent Protein-I-SceI fused to a glucocorticoid receptor ligand-binding domain. Upon induction with ligand, the I-SceI fusion moves to the nucleus and creates a single NHEJ focus, marked with γ -H2AX and 53BP1. The NHEJ focus coincides with the locus of CFP-lac repressor binding. (B) Similar experiment in cells expressing CFP-lac repressor and YFP tet repressor. Sequences on either side of the break are positionally stable on a time scale of minutes. Data are from ref [47].

Figure 4. A gene correction machine

The target cell contains a defective gene (indicated by a red X). Circular DNA is introduced as a nucleoprotein complex with a mutant Rag protein. Rag introduces a single strand nick and guides the DNA into the homologous recombination pathway of DNA repair. There is no potential for random insertion, and the defective gene is eliminated at the end of the process. The gene correction machine is depicted as encapsulated in a nanoparticle that aids in targeting to the correct intracellular compartment (see text).

Table 1

Major protein components of the NHEJ complex in mammalian cells

Function	Complex	Subunit	Comments
Core NHEJ	Ku	Ku70	Initial end recognition
		Ku80	Initial end recognition
	DNA-PKcs	-	Sequesters ends, promotes synapsis
	DNA ligase	DNA lig IV	Phosphodiester bond formation
		XRCC4	Accessory to DNA ligase IV
		XLF	Accessory to DNA ligase IV
Processing	M•R•N	Mre11 Rad50 NBS1	Promotes synapsis, has intrinsic nuclease, cell cycle regulatory activities. Requirement defined genetically in budding yeast only
		Artemis	Endo- and exonuclease
		Pol X family	DNA polymerases, fill gaps
		PNK	Polynucleotide kinase, adds phosphate to DNA
	Repair foci		
Repair foci	γ-H2AX	-	Phosphorylated histone variant
	ATM	-	Phosphorylates H2AX and other substrates
	53BP1	-	Cell-cycle regulatory activity
	MDC1	-	Cell-cycle regulatory activity
	HDAC4	-	Histone deacetylase

X-ray crystallographic structures are available for many of the core NHEJ components, which will facilitate efforts to determine higher-order organization on the nanometer scale.

Table 2

Methods for site-specific labeling of proteins [71]

Method	Tag Size ^a	Examples	Restrictions	References
Genetic fusion with fluorescent protein				
GFP derivatives	238	Enhanced green, yellow, cyan fluorescent proteins	Bleaching, size	[35,36]
DsRed derivatives	225	DsRed, mCherry, mPlum	Bleaching, size	[72–74]
Site-specific attachment of exogenous fluorophore				
Peptide tags	6–38	biotin acceptor, oligohistidine, tetracysteine, dye-binding “fluorettes” dihydrofolate reductase;	lack of specificity, toxicity	[75–79]
Enzyme or receptor domains	77–207	FKP12 (binds FK506 antibiotic); alkylguanine transferases, acyl carrier protein	endogenous background	[37,38,80–84]
Intein/native chemical ligation	150 ^a	Portion of intein (protein)	Slow, large ligand	

^aThe tag is temporary and is cleaved inside the cell.