UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Dynamics of nucleotide excision repair complex assembly and disassembly in vivo

Luijsterburg, M.S.

Publication date
2008

Link to publication

Citation for published version (APA):
Luijsterburg, M. S. (2008). *Dynamics of nucleotide excision repair complex assembly and disassembly in vivo*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Mar 2023

Chapter 8

Perspective: Assembly of Multi-Protein Complexes that Control Genome Function

Christoffel Dinant, Martijn S. Luijsterburg, Wim Vermeulen, Adriaan B. Houtsmuller and Roel van Driel

In preparation

Introduction

The nucleus is the information repository of the cell. In this ~300 μm^3 container, mammalian cells store the genetic material. The 3 billion base-pairs of the human genome associate with histone proteins to form a chain of basic units of chromatin: the nucleosomes. Functioning of the genome is controlled by a variety of processes that each involve tens of proteins that cooperate in space and time. Examples are transcription initiation complexes, replication machineries and DNA repair mechanisms. In vivo studies have revealed that many proteins that control genome function rapidly diffuse inside the mammalian nucleus (Phair and Misteli, 2000) with apparent diffusion rates ranging between ~1 and 10 $\mu m^2/s$ depending on the shape and the size of the molecule. If such a diffusing protein encounters a site for which it has a finite affinity, it binds for a certain time depending on its dissociation rate. Many nuclear proteins exchange fast between the bound and free state at the scale of seconds (e.g. transcription factors) or minutes (e.g. repair proteins; Luijsterburg et al., 2007; Phair et al., 2004). Due to the short binding time of individual chromatin-binding proteins, the formation of a multiprotein complex consisting of tens of protein molecules is a low probability event, unless special processes are implemented. Although the binding kinetics of many individual proteins have been measured, little is known about how proteins assemble into the functional multiprotein complexes that are involved in controlling genome function in living cells. In this perspective, we focus on the general mechanism and kinetics of the assembly of multiprotein complexes involved in genome functioning, studied using live-cell imaging. We discuss how proteins find their target site on the genome, a process that has mainly been studied using bacterial proteins binding to naked DNA. Recent in vivo studies suggest that similar mechanisms are used by chromatin-binding proteins in eukaryotes. Additionally, we give an overview of the binding kinetics of proteins involved in genome controlling processes, such as transcription of rRNA and mRNA genes, replication of the genome and DNA repair. Finally, we discuss how live-cell studies aided by mathematical modelling have unveiled characteristic properties of assembly of multi-protein complexes on the chromatin fibre, which appears to be similar for different genome controlling processes. We argue that complex data sets from kinetic studies cannot be interpreted intuitively and require mathematical modelling to gain comprehensive insight into the dynamic organization and functioning of multi-protein complexes that control genome function.

How do site-specific proteins find target site on chromatin?

Essentially all processes that control genome function are carried out by complexes that contain tens of proteins that assemble on specific genomic sites to exert their function. Assembly of such multi-protein complexes at specific sites is often initiated by a recognition protein that is able to, for example, discriminate between damaged and

non-damaged DNA (initiating DNA repair) or recognize a specific sequence in a promoter region (initiating transcription). Several mechanisms have evolved to ensure that site-specific proteins bind their target sites with sufficient rates. Proteins often bind to specific and non-specific sites with similar on-rates. Differences in affinity are often determined by a slower rate of dissociation (Hopfield, 1974). Site-specific proteins diffuse through the nucleus until they transiently interact with chromatin by chance. Since non-specific (i.e. low affinity) sites are usually present in large excess over specific sites, it is more likely that a protein binds a non-specific site (Misteli, 2008). Thus, once a protein interacts with a low affinity site on chromatin it will be transiently bound and then dissociate (since it has a higher dissociation rate from low affinity sites) until it encounters a specific site (i.e. high affinity) from which it dissociates more slowly. Multiple association/dissociation events on non-specific sites will take place before a protein binds to a specific site more stably (Gorski et al., 2006; Misteli, 2007). Some proteins associate with their target sites several orders of magnitude faster than diffusion-limited association (~ 10^8 $M^{-1}s^{-1}$) to DNA in vitro (Berg et al., 1981; Riggs et al., 1970). This rapid association can be accounted for by movement of a protein from its initial non-specific site to its target site by one-dimensional (1D) diffusion along the DNA by a sliding and/or hopping mechanism, which involves electrostatic DNA-protein interactions (Berg et al., 1981; Elf et al., 2007; Halford and Marko, 2004; Shimamoto, 1999). It is not likely that many proteins bind to DNA with association rates above the diffusion limit in vivo due to the high concentration of salt (equivalent to >150 mM NaCl) in cells, which reduces electrostatic DNA-protein interactions and thus one-dimensional diffusion (Halford and Marko, 2004). Sliding implies diffusion along the DNA helix without losing contact, whereas hopping involves dissociation and re-binding near to the site where the protein last dissociated. Hopping may occasionally occur to a more distant site on linear DNA, since the genome is present in a highly folded state (Halford and Marko, 2004). However, experiments show that hopping proteins tend to associate close to the site where they last dissociated. It may be that this is due to long-range electrostatic DNA-protein interactions that "pull" the protein back to the DNA (Halford and Marko, 2004).

Many DNA-binding proteins are able to move along a DNA strand without dissociating from it, including restriction enzymes, transcription factors and DNA repair proteins (Blainey et al., 2006; Coppey et al., 2004; Gowers and Halford, 2003; Gowers et al., 2005; Graneli et al., 2006; Harada et al., 1999; Shimamoto, 1999). However, at physiological ionic strength, proteins only diffuse along the DNA over distances of about 50 bp, suggesting that 1D diffusion is not the main mode of translocation of site-specific proteins (Gowers et al., 2005). Several examples indicate that site-specific proteins in eukaryotes also have affinity for non-specific DNA sites. For instance, repair factor XPC continuously associates with and dissociates from chromatin and

occasionally encounters a lesion that it binds to with higher affinity. More precisely, at each moment 55% of the XPC molecules are bound to DNA for on average 300 ms in undamaged cells (Hoogstraten and Vermeulen, in press). These 300 ms likely represent the time that XPC diffuses along the DNA until it dissociates or encounters a DNA lesion. Other damage recognition proteins, Rad51 and hOgg1, were shown by single molecule techniques to diffuse along double stranded DNA and bind with higher affinity upon encountering a lesion (Blainey et al., 2006; Graneli et al., 2006). Similarly, proteins involved in transcription initiation were found to move along DNA by 1D diffusion and bind more tightly once encountering promoter regions or other regulatory elements (Elf et al., 2007; Harada et al., 1999). Probably, proteins alternate between 1D sliding over short distances and short hops near the dissociation site (1D hop) interrupted by periods of free diffusion. In eukaryotes, sliding is likely restricted to the linker regions that have no histones bound (Hannon et al., 1986). In agreement, promoter localization by RNA polymerase is equally efficient for promoters in the linker DNA of chromatin in living cells as for promoters in naked DNA (Hannon et al., 1986). Moreover, the C-terminus of p53 was shown to mediate 1D-diffision along DNA and a p53 mutant lacking this domain was unable to bind specific promoters in vivo (McKinney et al., 2004), suggesting that 1D-diffusion is also important for locating target sites in a chromatin context. Since sliding is mainly due to electrostatic interactions, the sliding properties will be determined by the distribution of (mainly positively) charged residues on the protein surface that interact with DNA. A protein that displays 1D diffusion is often envisioned to track the major groove of DNA, thus spiralling around the helix as it diffuses along the DNA. Another possibility is that proteins diffuse freely on the DNA surface (termed 2D-diffusion). Experiments revealed that such a mechanism is indeed employed by bacterial proteins and modelling showed that 2D diffusion would allow a protein to bypass obstacles such as nucleosomes (Kampmann, 2004). Therefore, 2D diffusion of proteins might be more relevant in a chromatin context. Mere 3D diffusion is not sufficient to explain the high rates at which proteins associate with target sites (Halford and Marko, 2004). Mechanism that increase the rate of binding are diffusion along the DNA (either tracking the major groove or by diffusing freely on the cylindrical DNA surface) for short stretches combined with “hopping” and “jumping”. Clearly, the contribution of these processes to finding the target site depends on the biophysical properties of the protein and the number and nature of the binding sites.

Assembly of multi-protein complexes that control genome function

Finding of a target site by a recognition-protein is only the starting point for genome controlling processes. Upon recognition of an origin of replication, DNA lesion or promoter by a recognition protein, multi-protein complexes often containing tens of proteins are assembled that subsequently carry out replication, DNA repair or transcription. Here, we give an overview of live-cell imaging studies on proteins that

form such multi-protein complexes during DNA repair, transcription and replication. First, we discuss the mechanism of protein complex assembly during DNA repair by the nucleotide excision repair system. This process has been studied extensively and may therefore serve as a paradigm for other chromatin-associated processes.

Assembly of DNA repair complexes.

To protect the integrity of the genome, multiple DNA repair mechanisms have evolved to deal with specific DNA injuries (Essers et al., 2006; Hoeijmakers, 2001). For example, nucleotide excision repair (NER) removes helix-distorting injuries that affect one of the DNA strands, whereas homologous recombination (HR) and non-homologous end-joining (NHEJ) repair double strand breaks (DSBs). Particularly the kinetics of proteins involved in NER have been well studied. NER involves the assembly of repair complexes that contain up to 30 polypeptides, which cooperate in space and time. Damage-recognition protein XPC was recently shown to continuously associate and dissociate with chromatin for on average 300 ms per binding event (Hoogstraten and Vermeulen, in press) At any moment, about half of the XPC molecules is bound to chromatin and half diffuses freely. It is likely that the chromatin-bound fraction of XPC diffuses along the DNA during this short period, similar to 1D diffusion reported for some restriction enzymes and the lac repressor (Elf et al., 2007). In the absence of damage, other NER proteins XPA, XPG, and ERCC1/XPF diffuse rapidly through the nucleus with rates that correspond to their molecular size (Houtsmuller et al., 1999; Luijsterburg et al., 2007; Rademakers et al., 2003; Zotter et al., 2006). Upon DNA damage induction by UV-C light, XPC occasionally encounters a helix-distorting lesion to which it binds more stably ($t_{1/2}$ = 25s) (Hoogstraten et al., 2008). Lesion detection by XPC subsequently triggers assembly of NER complexes on the site of damage from freely diffusing constituents. NER proteins XPG, TFIIH and ERCC1/XPF rapidly exchange ($t_{1/2}$ ~1 min) with emerging repair complexes, while XPA exchanges somewhat slower (~2 min) and, as mentioned above, XPC exchanges is somewhat faster ($t_{1/2}$ = 25s) (Houtsmuller et al., 1999; Luijsterburg et al., 2007; Rademakers et al., 2003; Zotter et al., 2006). Fast exchange of proteins between the diffusing and bound state may ensure a high free concentration of these factors allowing a dynamic competition for binding sites, which may facilitate cross-talk between different processes that control genome function. However, this fast exchange also makes the formation of functional multi-protein repair complexes a time-consuming process, since binding of all proteins to the same site at the same time has a low probability. Rather, there will be an ensemble of mostly in complete assembled repair complexes. Only a few complexes will have all the repair proteins bound at the same time. This does not appear specific for DNA repair complexes, since protein complex assembly was also shown to be slow during other genome-controlling processes (see below and (Darzacq et al., 2007; Dundr et al., 2002)). Indeed, mathematical modeling suggests that by far most of the repair time is

spent on assembly of functional complexes, which is an inherent property of large multi-protein complexes (Luijsterburg and van Driel, unpublished). NER involves some of the same enzymatic reactions (e.g. helix unwinding and DNA synthesis) that are carried out during transcription and replication, which involves proteins that are shared between these different processes, such as TFIIH, RPA and PCNA. The nucleoplasmic pool of TFIIH is in a continuous equilibrium between RNA pol I, RNA pol II transcription and NER (if DNA damage is present). The binding time of TFIIH at transcription sites is in the order of seconds, while TFIIH is bound much longer to NER intermediates during repair ($t_{1/2}$ ~50 s) (Hoogstraten et al., 2002). Similarly, PCNA and RPA are shared factors between replication and several repair pathways including NER (Essers et al., 2005; Solomon et al., 2004). PCNA is bound for about 3 minutes during replication, while its exchange rate at sites of DNA damage is almost twice as long ($t_{1/2}$ ~300s). RPA is also more stably associated with repair sites than with replication (Solomon et al., 2004). Thus, it appears that TFIIH, RPA and PCNA have higher affinities for DNA repair intermediates than for sites of transcription and replication, respectively.

Repair of DSBs by HR involves assembly of a protein complex that includes the formation of a Rad 51 filament (in which the actual recombination events takes place) and binding of chromatin remodelling factors (Rad54) and additional repair proteins such as Rad52 and RPA. Live cell imaging revealed that Rad51 filaments are highly stable and the residence time of Rad51 proteins is in the order of hours. Exchange rates of Rad52 (~1min) and Rad54 (~10 s) are much faster (Essers et al., 2002). Thus, it appears that Rad51 is a structural component during HR, which serves as a binding platform for several other repair proteins that exchange rapidly between the bound and free diffusing state. Such a mechanism is reminiscent of replication in which PCNA is bound relatively stably. Besides HR, DSBs can be repaired by NHEJ if a sister chromatid is not present (i.e. such as in G1), which involves, among others, the ring-shaped Ku70/80 dimer and the kinase DNA-PKcs. Current models suggest that multiple Ku70/80 dimers encircle the broken DNA ends and become trapped there. After ligation of the broken ends, it has been suggested that Ku70/80 could be permanently trapped (Mari et al., 2006). However, live-cell imaging challenged this model and revealed that binding of Ku70/80 to broken DNA ends is reversible and that exchange between bound and soluble pools occurs within 40s. Moreover, Ku70 was shown to recruit LigIV via XRCC4 to broken DNA ends, which in turn joins the broken ends together (Mari et al., 2006). Similar to Ku70, DNA-PKcs was shown to exchange between soluble and DNA bound pools within 1 min. Nevertheless, when DNA-PKcs cannot be phosphorylated or perform its kinase activity, a much larger fraction of the DNA-PKcs pool is bound for longer times (Mari et al., 2006; Uematsu et al., 2007). These studies highlight the importance of post-translational modification of repair proteins and show that, in case of DNA-PKcs, phosporylation is needed to

lower its affinity for repair intermediates.

Collectively these studies show that, with the exception of Rad51 filaments, repair machineries assemble at sites of DNA damage from freely diffusing components and form short-lived complexes on damaged chromatin. Affinity differences between specific and non-specific sites are small. Likely, the affinity of repair proteins is precisely tuned such that transient binding is sufficient to assemble complexes at specific (in this case damaged) sites with an acceptable rate, while at the same the low affinity ensures that complex assembly at non-specific sites is not very likely. Indeed, a recent study showed that tethering of DSB repair proteins Mre11, Rad50 or Nbs1 to chromatin elicits a DNA damage response including activation of Chk1/Chk2 and cell cycle arrest (Soutoglou and Misteli, 2008). This confirms that binding of a single repair protein with a low koff (high affinity binding) is sufficient to trigger a cellular response.

Assembly of transcription initiation and elongation complexes.
Transcription involves assembly of a multi-protein transcription initiation complex on the chromatin fibre. Various live-cell imaging studies in combination with kinetic modelling have unveiled that, like DNA repair proteins, transcription factors, co-activators and RNA polymerases (RNA pol) bind rapidly and reversibly to target-sites (promoters) in a stochastic fashion (Becker et al., 2002; Hager et al., 2006). Occasionally, these factors assemble in a way that leads to onset of transcription and the production of RNA (Darzacq et al., 2007). Several transcription factors and co-activators (e.g. GR, GRIP-1, p53, TFIIB and TFIIH) diffuse rapidly through the nucleus and at any given time about 15% - 25% of these proteins is bound for 3-5 s to chromatin (Chen et al., 2002; Hoogstraten et al., 2002; Mueller et al., 2008). Although short residence times (i.e. several seconds) are common for transcription factors, some have binding times in the order of 1 min (e.g. AR and TBP) (Chen et al., 2002; Farla et al., 2004). The majority (~85%) of RNA pol II in living cells rapidly exchanges within on average 6 s, reflecting reversible binding to promoters or to non-specific sites with low affinity (Darzacq et al., 2007). The remaining 15% of RNA pol II molecules binds for on average about 1 min, reflecting transcription initiation attempts at the promoter. About half of these 1 min binding events leads to elongation with ~70 nt/s during which RNA pol II exchanges slowly from transcribed genes (~10 min) (Darzacq et al., 2007; Kimura et al., 2002). These live-cell studies combined with mathematical modelling reveal that only 1% of the RNA pol II binding events at promoters lead to a complete RNA molecule, showing that the majority of RNA pol II-promoter interactions are not productive (Darzacq et al., 2007). This at first sight inefficient onset of transcription may indicate that there is only a low probability of forming an active transcription initiation complex at a promoter. Intriguingly, a small fraction (~5%) of elongating polymerases pauses for several minutes without dissociating.

Thus, at any time, at least one-fourth of transcribed genes have a paused RNA pol II bound resulting in discontinuous production of mRNAs (Darzacq et al., 2007), providing an explanation for transcriptional bursts. Recently, the steady-state level of a bound transcription factor to a natural promoter was found to oscillate in vivo on a time-scale of 20-40 min after transcription initiation by stimulation with a hormone, while the exchange rate of this transcription factor with chromatin was in the order of 1 min. These slow oscillations in steady-state levels of bound transcription components have also been detected by ChIP (Metivier et al., 2006) and it has been suggested that these slow cycles reflect stable interactions of activators with promoters leading to mRNA production, while fast cycles reflect transient interactions of activators that do not result in mRNA production (Metivier et al., 2006).

Live-cell imaging suggested that rapid exchange (within 1 min) of transcription factors reflects active transcription, whereas slow cycles reflect gradual changes in accessibility of binding sites at promoters (resulting in 20-40 min oscillations) (Karpova et al., 2008). In agreement, the occupancy of histones at promoters was also found to oscillate and to inversely mirror the slow cycle of a transcriptional activator. However, it is currently not clear, why the number of accessible binding sites of transcription factors gradually changes in time and whether this is functionally important for transcription regulation. In addition to RNA pol II, transcription of rRNA genes by the RNA pol I system is also highly dynamic (Dundr et al., 2002). The majority of pre-initiation factors (UBF1 and 2), transcription factors (TAFI48) as well as individual subunits of RNA pol I rapidly exchanges within 5 s at rRNA genes, while TFIIH exchange in nucleoli is considerably slower (~25s) (Hoogstraten et al., 2002). Similar to RNA pol II transcription, only 1-3% of the RNA pol I binding events result in elongation, which is inefficient in terms of association/dissociation steps needed to initiate transcription. However, because there are several hundred transcription factors binding events per second, such an 'inefficient' mechanism still results in rapid assembly of functional complexes in time. Interestingly, RNA pol I subunits exchange on rRNA promoters about 4 times slower in S-phase, during which the rRNA transcriptional output is much higher than in G1, suggesting that longer binding time of individual RNA pol I subunits to promoters results in more efficient assembly of functional RNA pol I complexes able to produce an rRNA (Gorski et al., 2008). Thus, modulation of RNA pol I assembly kinetics is a mechanism to control the transcriptional output of rRNA genes in living cells. In conclusion, these studies show that the assembly of active transcription initiation complexes is slow, similar to the formation of repair complexes. It is not likely that tens of proteins bind to the same site at the same time and the onset of transcription therefore takes multiple association/dissociation events. Additionally, the 'inefficiency' of complex assemble during transcription serves as a regulatory mechanism that reduces the incorporation of wrong (i.e. non-specific binding) proteins in the complex, a process named kinetic

proofreading (see below).

Assembly of replication complexes.
Faithful duplication of the genome is essential to maintain genome integrity and cellular identity. Besides DNA polymerases δ and ε, many proteins are involved in replication, including clamp-loader PCNA, endonuclease Fen1, ssDNA-binding protein RPA and DNA ligase 1 (Lig1) (Gorski et al., 2008; Liu et al., 2004; McCulloch and Kunkel, 2008). The leading strand is synthesized continuously during replication, while lagging strand synthesis requires the discontinuous synthesis and joining of Okazaki fragments. Fen1, RPA and Lig1 (and probably DNA polymerase δ), are specifically required for synthesis of the lagging strand (McCulloch and Kunkel, 2008). Initiation of replication involves origin recognition complex (ORC) proteins, which bind chromatin and exchange rapidly with soluble pools within 2 min (McNairn et al., 2005). This transient binding of ORC proteins to chromatin indicates that the mechanism to assemble these replication initiation proteins likely involves sliding and hopping, similar to assembly of repair and transcription initiation complexes. Replication of DNA occurs in discrete areas in the nucleus termed replication foci or factories, which appear at the onset of S phase and disappear as soon as replication is finished (Leonhardt et al., 2000). These microscopically visible replication foci represent clusters of replication forks and associated replication machineries and are often several MDa in size. Although replication foci appear to be static structures, the proteins that make up these structures exchange rapidly between the bound and freely diffusing pool (Kitamura et al., 2006; Misteli, 2007).

Live-cell imaging has shown that there is a remarkable difference in exchange kinetics between different replication proteins to replication foci. For example, exchange of clamp-loader PCNA at replication foci takes 3 min on average (Essers et al., 2005; Solovjeva et al., 2005; Sporbert et al., 2002), while proteins involved in lagging strand synthesis exchange much faster compared to PCNA. For instance, Fen 1 and DNA lig 1 exchange at replication foci takes about 1s (Solovjeva et al., 2005; Sporbert et al., 2005). Similarly, RPA exchanges in seconds (redistribution in 1 min) (Sporbert et al., 2002). The exchange rates of these proteins are in the same range as the time it takes to synthesize one Okazaki fragment (estimated at 6 s), which is roughly 180 bp in size (Jackson and Pombo, 1998; Solovjeva et al., 2005; Sporbert et al., 2005), suggesting that these proteins bind during the synthesis of one Okazaki fragment and dissociates upon completion of such a fragment. Fen1 and DNA lig1 are indeed required for Okazaki fragment maturation, performing removal of the RNA-DNA primer and sealing of the remaining nick, respectively (Liu et al., 2004). This favours a scenario in which proteins such as RPA, Fen1 and DNA lig1 exchange after each Okazaki fragment has been synthesized, whereas PCNA remains bound for the synthesis of multiple Okazaki fragments (Sporbert et al., 2005). Thus, it seems that PCNA serves

as a binding platform for many proteins that are transiently involved in replication, including proteins involved in epigenetic maintenance such as Dnmt1, which methylates the newly synthesized DNA during replication (Schermelleh et al., 2007). The use of a binding platform for transiently interacting proteins is conceptually similar to transcription elongation, during which RNA pol II exchanges very slowly, while other proteins, such as TFIIB or TFIIH, can transiently interact with the active transcription machinery (Darzacq et al., 2007). Whether DNA polymerases act similarly as RNA pol II with respect to chromatin association kinetics has not been studied so far.

Understanding assembly and functioning of genome- controlling complexes. The study of GFP-tagged proteins involved in chromatin processes generates large and complex sets of data that generally cannot be interpreted intuitively (Phair and Misteli, 2001). Kinetic modelling of quantitative in vivo data is a powerful tool in obtaining mechanistic insight into genome-controlling processes. It allows estimation of biophysical properties of proteins such as diffusion, association and dissociation rate constants that cannot be determined directly in vivo. Moreover, modelling of the kinetic properties of chromatin-associated systems provides detailed insight into their behaviour (Darzacq et al., 2007; Dundr et al., 2002; Politi et al., 2005). Luijsterburg and van Driel, unpublised results). Although it is the often-held view that multi-protein complexes are assembled in a stepwise fashion, this is not supported by several in vivo studies. These studies combined with mathematical modelling revealed that proteins only occasionally form the complete protein complex on chromatin. For example, components of the RNA pol I machinery frequently associate with ribosomal genes with a rate of several hundred molecules per second, but only occasionally (1-10% of the binding events) form the correct complex that is capable of transcription (Dundr et al., 2002). Likewise, in vivo studies and modelling revealed that only 1% of the RNA pol II-promoter interactions results in completion of an mRNA (Darzacq et al., 2007), suggesting the formation of the complete protein complex is a low probability event. Recent modelling of the kinetics of NER in terms of the collective action of eight repair proteins and six enzymatic steps, revealed that protein complex assembly during NER is remarkably slow compared to the enzymatic processes. More than 99% of the repair time is spent on the assembly of catalytically active complexes, whereas enzymatic reactions during repair (such as unwinding of DNA and dual incision) take only a small fraction of the total repair time (Luijsterburg and van Driel, unpublished). Although the term inefficient has been used for these processes, it should be noted that if several hundred molecules bind per second and even only 1% of the binding events results in assembly of a functional complex, this still results in complex formation on specific sites at sufficient high rates. In conclusion, a low probability to assemble the correct protein complex appears a shared characteristic of different genome controlling processes.

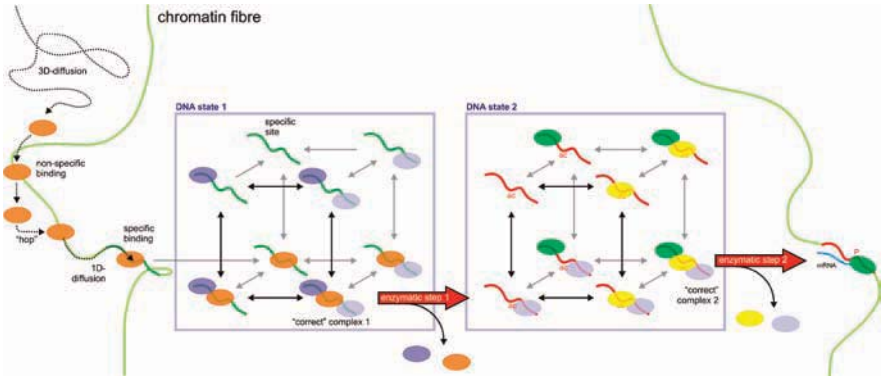

Figure 1. Model for binding of a site-specific protein to a target site and subsequent assembly of a multi-protein complex on that site. A site-specific protein (orange oval) diffuses rapidly inside the nucleus and binds non-specifically to chromatin (represented by the light-green line), dissociates and re-binds close to the site it dissociated from (so-called "hopping"). Finally, it moves along chromatin by 1-dimensional diffusion and encounters a specific site (dark-green) to which the protein binds more stably. The orange protein mediates the assembly of a complex consisting of two additional proteins (light-purple and dark-purple). Binding of all these proteins is stochastic and 8 different assembly states can be formed on the specific site consisting of one or a combination of the 3 proteins or the site can be devoid of any protein. Once the "correct" complex containing all 3 proteins is formed, the specific site is modified (for example acetylated shown in red) resulting in dissociation of the orange and dark-purple protein while the light-purple protein remains bound (since it has affinity for the altered state while the other proteins do not). The red arrow reflects an enzymatic step (in this case acetylation, which is in principal irreversible). This altered state is the substrate for a new set of proteins (the green and yellow protein and the light-purple protein from the last box) to bind to. Complex assembly is again stochastic and 8 different assembly states can be formed. Assembly of the "correct" complex containing all three proteins in the second box results in an enzymatic step that produces mRNA, resulting in dissociation of the yellow and light-purple protein. The probability of the overall reaction (i.e. binding of 5 different proteins to the same site) is increased by splitting the reactions in assembly of complex 1 and complex 2 separated by an irreversible enzymatic reaction. The irreversible step drives the reaction forward. Completion of processes involving more proteins can be kinetically driven by multiple irreversible reactions.

In this scenario, proteins bind stochastically (i.e. not in a fixed order of binding) to specific sites on chromatin to form an ensemble of protein complexes with different subunit composition. The correct protein complex, capable of performing a specific function (repair, transcription, replication), is formed with a low probability. This "chaotic" view on complex assembly is radically different from the often-held view that complex assembly on chromatin occurs through an ordered and stepwise mechanism in which each protein is incorporated in a long-lived chromatin-bound complex. This stochastic mechanism of complex assembly is schematically depicted in figure 1. Completion of a process in which assembly of large protein complexes is a low probability event can be kinetically driven by irreversible (often-enzymatic) reactions, which split up the process in smaller sub-processes that are executed sequentially. Such a mechanism, known as kinetic proofreading, increases the specificity of molecular processes above the level of difference in free energy between correct and incorrect

substrates (Hopfield, 1974). This creates a system that rejects 'wrong' proteins, despite the similar kon rates of specific and non-specific interactions. Many genome-controlling processes involve enzymatic reactions including ATP-hydrolysis, DNA unwinding, cleavage of DNA, DNA synthesis and ligation, which due to their often irreversible nature can drive genome controlling processes to completion (see fig 1). Moreover, these irreversible reactions allow the specificity of DNA-binding proteins for specific sites and the specificity at which multi-protein complex are formed to be increased at a level that is not present at the level of individual proteins. Thus, it appears that cells have evolved robust mechanisms to assemble multi-protein complexes at specific sites on chromatin with high accuracy. These mechanisms can only be unravelled using live-cell analyses combined with mathematical modelling and this combined approach is expected to provide insight into behaviour of systems that control the genome, in the years to come.

Acknowledgement

This work was supported by ZonMW, the Netherlands (grant 912-03-012).