large FHA domain family. For example, the amino acid residue in the FHA2 domain corresponding to Asn 86 in the FHA1 domain is an arginine. Clearly, more structural studies are necessary to determine whether this arginine residue plays an important role in phosphothreonine recognition by direct binding to the phosphate group. Such structural analyses will also help determine whether FHA domains are simply phosphothreonine-specific protein binding domains, or if some of them may possess other biochemical functions such as phosphoserine recognition.

From studies of SH2 and PTB domains, we have learned that basic molecular mechanisms of fundamental importance such as phosphotyrosine binding can be used in different contexts in a wide variety of biological processes including mitogenic receptor signal transduction. Our new understanding of phosphothreonine recognition provided by Durocher *et al.*[5] will undoubtedly facilitate our understanding of the molecular mechanisms by which FHA domain-containing proteins function in nuclear processes, including DNA repair and replication. If the FHA domains are indeed as versatile as SH2 and PTB domains, exciting things are bound to emerge.

Acknowledgment
I thank S.P. Jackson and S.J. Smerdon for providing the FHA1 domain coordinates, and A. Farooq and K.S. Yan for helping with figure preparation and discussion.

Ming-Ming Zhou is in the Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, 1425 Madison Avenue, Box 1677, New York, New York 10029, USA. email: zhoum@inka.mssm.edu

1. Pawson, T. & Nash, P. *Genes Dev.* **14**, 1027–1047 (2000).
2. Sun, Z., Hsiao, J., Fay, D.S. & Stern, D.F. *Science* **281**, 272–274 (1998).
3. Li, J., Smith, G.P. & Walker, J.C. *Proc. Natl. Acad. Sci. USA* **96**, 7821–7826 (1999).
4. Durocher, D., Henckel, J., Fersht, A.R. & Jackson, S.P. *Mol. Cell* **4**, 387–394 (1999).
5. Durocher, D. *et al. Mol. Cell* **6**, 1169–1182 (2000).
6. Hofmann, K. & Bucher, P. *Trends Biochem. Sci.* **20**, 347–349 (1995).
7. Bell, D.W. & Davies, J.E. *Trends Cell Biol.* **10**, 32–38 (2000).
8. Carney, J.P. *et al. Cell* **93**, 477–486 (1998).
9. Matsuoka, S., Huang, M. & Elledge, S.J. *Science* **282**, 1893–1897 (1998).
10. Eck, M.J., Shoelson, S.E. & Harrison, S.C. *Nature* **362**, 87–91 (1993).
11. Waksman, G. *et al. Nature* **358**, 646–653 (1992).
12. Zhou, M.-M. *et al. Nature* **378**, 584–592 (1995).
13. Zhou, M.-M. *et al. Nature Struct. Biol.* **3**, 388–393 (1996).
14. Eck, M.J., Dhe-pagnon, S., Trüb, T., Nolte, R. & Shoelson, S.E. *Cell* **85**, 695–705 (1996).
15. Shi, Y., Hata, A., Lo, R.S., Massague, J. & Pavletich, N.P. *Nature* **388**, 87–93 (1997).
16. Liao, H., Byeon, I.-J.L. & Tsai, M.-D. *J. Mol. Biol.* **294**, 1041–1049 (1999).
17. Yaffe, M.B. *et al. Cell* **91**, 961–971 (1997).
18. Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T. & Noel, J.P. *Nature Struct. Biol.* **7**, 639–643 (2000).

Some assembly required

Stephen C. Kowalczykowski

Many biological processes require the ordered assembly and disassembly of complex structures. By using common structural elements, conformational switching and competitive interactions, the proper temporal assembly of the structures is assured, while simultaneously taking full advantage of combinatorial flexibility.

It's been said that the words 'some assembly required' strike fear in the hearts of many otherwise competent adults. The same may also apply to some biologists who would prefer that their favorite biological machines come pre-assembled and ready for action. However, in some cases, cellular realities require assembly (and disassembly) of essential cellular structures. Recent evidence (see below) suggests that this is true for the protein–DNA complexes that constitute the functional entities involved in DNA replication, recombination, and repair. This notion is strengthened by the recent findings of Mer *et al.*[1] demonstrating that the eukaryotic single-stranded DNA-binding protein, called replication protein-A (RPA), contributes to the orchestration of sequential events in DNA recombination and repair *via* a structurally common set of protein–protein interactions that involve protein components from each of these pathways.

According to the model they propose, RPA coordinates the cellular response to DNA damage by directly recruiting needed components of the repair machinery to sites of DNA damage (in conjunction with other protein specificity factors). Then, as a consequence of these mutually competitive interactions, they propose that the exchange or 'handing-off' of components of the repair machinery ensues[2] so that later steps of the repair process can progress.

The single-stranded DNA binding (SSB) proteins are ubiquitous. They participate in nearly every aspect of DNA maintenance and, not surprisingly, they are essential for cellular function[3]. Yet, they themselves have no enzymatic activity, and their 'only' function is to bind tenaciously to single-stranded DNA (ssDNA). However, since ssDNA is a common intermediate of DNA replication, recombination and repair, this capacity is essential, if for no other reason, than to protect the ssDNA from further untargeted modification (for example, nucleolytic degradation).

The first SSB proteins were identified in bacteria and phage: the eponymous SSB protein from *Escherichia coli* and T4-phage gene 32 protein. Their ability to bind tightly and cooperatively to ssDNA is a relatively unique hallmark, but protein–protein interactions between these proteins and other protein components of DNA replication and recombination are also vital. Subsequently, analogs were found in Eucarya and Archaea; however, the subunit composition of these proteins is different from that of the bacterial and phage prototypes. Rather than consisting of a single polypeptide, the eukaryotic proteins are heterotrimeric[4], and the archaeal proteins are either dimeric or monomeric[5,6]. Yet despite the structural differences, the behavior of each protein is similar — each binds ssDNA tightly, and each is devoid of enzymatic activity.

Human RPA interacts with many proteins involved in DNA replication,

recombination and repair[4]. The list is extensive and includes a component of the base excision repair system (BER), uracil-DNA glycosylase (UNG2); a component of the nucleotide excision repair system (NER), the xeroderma pigmentosum damage-recognition protein (XPA); and a member of the double-strand DNA break repair machinery (DSBR), a DNA annealing and recombination mediator protein (Rad52).

Given the many processes in which RPA participates, questions arise as to how it is recruited to these different types of DNA damage, and how do the subsequent steps of each repair process proceed? The naïve answer is that RPA binds ssDNA and, since ssDNA is the common intermediate for each process, ssDNA is the common denominator for RPA participation. However, this simplistic answer ignores the extensive number of protein–protein interactions of which RPA partakes.

This question of RPA function can be recast in terms of two limiting views regarding the nature and dynamics of the structures that comprise each of these repair systems. Does RPA interact with a pre-assembled complex for each of the different repair reactions in which it functions as a permanent component of many different static, pre-assembled repair 'factories' that persist in the cell? Or, alternatively, is RPA a component of dynamic structures that are assembled on an 'as needed' basis to deal with the demands placed on the cell at any given time. In this second view, the common RPA–ssDNA intermediate helps to pass responsibility onto the next component(s) of the process *via* common sites that permit 'handing-off' of protein components.

Recent structural and biochemical data argue for the latter process as being dominant in these and other processes related to DNA maintenance[7]. Mer *et al.*[1] established that UNG2, XPA, and Rad52 each interacts with a common site on the RPA32-subunit of RPA, and each uses a common motif of amino acids within a helical conformation to mediate these interactions. The interaction domain within each of these proteins maintains a similar disposition of the positively charged amino acid residues that interact with acidic patches on RPA32; in the case of UNG2, this interaction domain is clearly distinct from and on the opposing side of the DNA binding site. The interacting region in RPA32 folds into a winged helix-loop-helix structure, and this module moves independently of the DNA binding domain. Thus, by definition, the interactions of these proteins with RPA are mutually exclusive and, therefore, competitive.

However, given the abundance of SSB proteins in the cell, one could argue that sufficient excess RPA exists to satisfy the need of any number of structurally stable '-osomes' that might be required. However, biochemical data tend to argue against this notion. First, the components of NER, BER, and DSBR are generally not isolated as stable multicomponent protein complexes[8,9], instead, at most, heterodimeric species are detected in some cases (for example, XPC–hHR23 or XPF–ERCC1). Second, the biochemical reactions are typified by a series of temporally distinct, sequential assembly processes that are followed by the coordinated exchange of an existing component of the complex for a new protein component, to generate a new co-complex with the distinct structural components and biochemical activities that are needed to promote the next step in the biochemical process.

This mechanism, where different proteins trade places in dynamically defined complexes, is also consistent with the wealth of two-hybrid results that consistently reveal multiple pair wise interactions between proteins; when summed over all observed pairwise interactions, such results are often over-interpreted as evidence for some type of extended holoenzyme complex instead of the sequential 'hand-off' mechanism.

The 'hand-off' mechanism seems to hold true for NER, where RPA interacts with XPF–ERCC1 and XPG, the damaged DNA is excised and then, after dissociation of the damaged DNA–protein recognition complexes, DNA polymerase is recruited to ssDNA–RPA complex to complete the repair process[8]. Similarly, in DSBR, RPA is the first protein to bind the ssDNA produced by processing of the dsDNA break[10]. Rad52 then assembles with this complex either to mediate a DNA annealing reaction[11,12], or to mediate an exchange of RPA for Rad51 to form the DNA strand exchange complex that is needed for the recombinational repair process[13–15]. Finally, clever photo bleaching-recovery experiments establish that the NER complex *in vivo* is dynamic, with the ERCC1–XPF complex moving freely throughout the nucleus at a rate expected for diffusion of the heterodimeric complex, rather than for a larger pre-assembled complex[16]. Thus, the dynamic nature of the protein–DNA complexes that serve as the intermediates in these processes illustrates how mutually exclusive, and competitive, reactions can be harnessed through temporally ordered formation and protein exchange to produce new nucleoprotein complexes that are kinetically competent for the subsequent step in the reaction mechanism.

So, how then, does RPA know with which protein to interact? One possible answer to this question is that the interaction of a DNA damage-specific recognition protein directs the initial pathway in which RPA will participate. For example, if the DNA damage is a thymine dimer, then either XPC–hHR23B or XPA–RPA recognizes the damage, distorting the DNA further, and allowing RPA to bind[8,9]; recruitment of the next protein in the cascade is defined by interaction with both the damage-recognition complex and RPA. Similarly, if the damage is dUTP, then UNG2 is recruited; subsequent interactions with RPA and UNG2 recruit a repair DNA polymerase.

In other cases, assembly may depend on timing — that is, the duration of time that the lesion exists. For example, in DSBR (or in the mechanistically similar repair of ssDNA gaps), how can the RPA protein 'know' whether the ssDNA that is present resulted from the processing of a DNA break or from normal lagging strand DNA synthesis? The answer to this question may simply be a matter of the kinetic lifetime of this 'intermediate': the ssDNA produced by normal lagging strand synthesis should not persist for more than a minute or so. Thus, although RPA will be bound to this ssDNA, it will not exist sufficiently long to permit recruitment of the recombination machinery because lagging strand synthesis will eliminate the ssDNA. Alternatively, if the ssDNA was produced as part of DSBR, then the resulting ssDNA will remain unrepaired and unreplicated, until Rad52, then Rad51, and so forth, are recruited to the longer-lived RPA–ssDNA complex. Finally, in the absence of sufficiently rapid repair, the RPA–DNA complex may then become too long-lived for cellular comfort, and it may then serve an alternative function, which is to maintain or generate a signal for cell cycle arrest until some form of repair is set in motion[17].

Why bother to constantly assemble and disassemble complexes that the cell needs frequently; why not just deliver them pre-

assembled? In a word, the answer is flexibility. The advantage of 'combinatorial' behavior is fully appreciated in both chemistry and biology. Given the huge and entirely unpredictable types of DNA damage that might be encountered (including unanticipated modern insults to DNA), the repair machinery must have the flexibility to respond to all types of DNA damage. In the event that the wrong DNA repair complex assembles on any type of DNA damage, the complex can disassemble and dissociate, allowing a different subset of repair enzymes to act on the lesion. The second feature of the flexibility argument is that intermediates of repair processes have the option to proceed down different end-pathways, depending on the state of the cell[18]. For example, the ssDNA produced as part of DSBR can be handled in several different ways: it can be annealed to a complementary sequence; it can invade a homologous duplex sequence as part of a recombination mechanism; or it can be repaired by a replicative pathway[19]. Biochemically, the path chosen depends on the nature of the complex that assembles after the RPA–ssDNA complex — Rad52 promotes annealing and Rad51 promotes DNA strand invasion. Finally, since the sites of DNA damage are randomly distributed around the chromosome, the repair machinery must have the capacity to find them. Pre-assembled factories, while efficient, simply cannot move as quickly as individual components that are assemble on site. Thus, the dynamic nature of biological structures facilitates their relocation.

In conclusion, despite the anxiety associated with the phrase, 'some assembly required', many processes in biology benefit from the flexibility associated with an ordered, temporal assembly and exchange/disassembly of structural and functional intermediates. By 'handing-off' components in a specific manner that is dictated by the structures of, and interactions among, early intermediates, the cell assures that later steps of complex processes can progress.

Stephen C. Kowalczykowski is in the Sections of Microbiology and of Molecular and Cellular Biology, and in the Center for Genetics and Development at the University of California, Davis, Davis, California 95616, USA. email: sckowalczykowski@ucdavis.edu

1. Mer, G. *et al*. *Cell* **103**, 449–456 (2000).
2. Parikh, S.S., Mol, C.D., Hosfield, D.J. & Tainer, J.A. *Curr. Opin. Struct. Biol.* **9**, 37–47 (1999).
3. Kornberg, A. & Baker, T.A. *DNA Replication* (W. H. Freeman and Co., New York; 1992).
4. Wold, M.S. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
5. Chedin, F., Seitz, E.M. & Kowalczykowski, S.C. *Trends Biochem. Sci.* **23**, 273–7 (1998).
6. Kelly, T.J., Simancek, P. & Brush, G.S. *Proc. Natl. Acad. Sci. USA* **95**, 14634–14639 (1998).
7. Yuzhakov, A., Kelman, Z., Hurwitz, J. & O'Donnell, M. *EMBO J.* **18**, 6189–6199 (1999).
8. Araujo, S.J. & Wood, R.D. *Mutat. Res.* **435**, 23–33 (1999). [published erratum appears in *Mutat. Res.* 2000 Mar 20;459: 171–172 (2000)].
9. Wakasugi, M. & Sancar, A. *J. Biol. Chem.* **274**, 18759–18768 (1999).
10. Gasior, S.L., Wong, A.K., Kora, Y., Shinohara, A. & Bishop, D.K. *Genes Dev.* **12**, 2208–2221 (1998).
11. Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S. & Ogawa, T. *Genes Cells* **3**, 145–156 (1998).
12. Sugiyama, T., New, J.H. & Kowalczykowski, S.C. *Proc. Natl. Acad. Sc.i USA* **95**, 6049–6054 (1998).
13. Sung, P. *J. Biol. Chem.* **272**, 28194–28197 (1997).
14. New, J.H., Sugiyama, T., Zaitseva, E. & Kowalczykowski, S.C. *Nature* **391**, 407–410 (1998).
15. Shinohara, A. & Ogawa, T. *Nature* **391**, 404–407 (1998).
16. Houtsmuller, A.B. *et al*. *Science* **284**, 958–961 (1999).
17. Lee, S.E. *et al*. *Cell* **94**, 399–409 (1998).
18. Otterlei, M. *et al*. *EMBO J* **18**, 3834–3844 (1999).
19. Paques, F. & Haber, J.E. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).

Going green

The green fluorescent protein (GFP) is one of those tools, like the polymerase chain reaction, that it is hard to imagine living without. Tagging proteins with GFP has revolutionized biology, making it possible to visualize particular cells in living organisms, such as *Caenorhabditis elegans*, and to watch molecules move during biological processes, such as cell division.

The development of GFP as a molecular tool has its roots in the study of bioluminescence, a phenomenon found most commonly in marine organisms such as jellyfish, corals, and dinoflagellates. Bioluminescence differs from fluorescence in one important aspect. In the former, light is produced by an inherent chemical reaction whereas in the latter, radiation of a particular wavelength is absorbed by a fluorescent molecule, and subsequently light of a different wavelength is emitted by that molecule. It was discovered many years ago that certain organisms use bioluminescent proteins and fluorescent proteins in tandem to emit light of a particular wavelength, although the biological advantage of emitting such light is not known.

In the early 1960s, Shimomura *et al.*[1] purified the first light-producing protein from the jellyfish *Aequoria victoria*. They found that this protein, known as aequorin, produced a blue light when activated by calcium — not the green light that was observed in the animals naturally. It was thought that another factor, perhaps a fluorescent molecule, might participate to create the green light. In support of this idea, in the 1950s it had been reported that *A. victoria* produces a greenish fluorescence when subjected to long-wavelength UV irradiation[2]. The situation was later clarified by a number of experiments on several related systems (reviewed in ref. 3). Aequorin is associated with GFP in the light-producing cells and energy transfer from activated aequorin to GFP results in the emission of green light.

It had been shown that the chromophore of GFP is formed from its primary amino acid sequence (see ref. 3 for a review). Nevertheless, it was unclear if GFP would be capable of fluorescing in heterologous cells, as it was thought that perhaps some jellyfish-specific enzyme might be required for assembly of the chromophore. Experiments using *Escherichia coli* and *C. elegans* settled this issue in 1994, showing that GFP expressed from a cDNA construct could indeed fluoresce in other organisms[4].

Research continues today on bioluminescent organisms, which have yielded other photoproteins and fluorescent molecules for use as biological tools. The structure of one such molecule, the red fluorescent protein known as DsRed, is described on page 1133 of this issue.

Tracy Smith

1. Shimomura, O., Johnson, F.H. & Saiga, Y. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).
2. Davenport, D. & Nicol, J.A.C. *Proc. R. Soc. B* **144**, 399–411 (1955).
3. Chalfie, M. *Photochem. and Photobiol.* **62**, 651–656 (1995).
4. Chalfie, M. *et al. Science* **263**, 802–805 (1994).