Repairing information lacunae: Coupling of transcription with repair

Purnima Bhargava

DNA is the substrate/template for the components of the three essential processes of the cell which result in the replication, transcription or repair of the genetic material. Recent evidences show that these fundamental biochemical processes necessary for the life and its propagation, are intricately and meticulously interwoven. A closer look reveals how the fine tuning of the three processes, viz. replication, repair and transcription of the DNA, through common knobs has helped the cells to generate and maintain mechanisms falling higher on the evolutionary scale and cope with the pressure of developing more control mechanisms. The genetic material of a cell is constantly subjected to a spectrum of events resulting in mutations, lesions or losses of information. Obviously, these deleterious effects cannot be tolerated by the cell for its survival. DNA repair has evolved as an essential mechanism to overcome this problem. Processes like replication and transcription, during which the DNA double helix is unwound, are almost invariably preceded, followed or accompanied by a careful scan of the DNA strands by the repair machinery to ensure that wrong information is not propagated. This is more important for the transcribed strand because vital information is copied from this strand.

It has been known for a long time that cells repair the active genes and the transcribed strand of DNA with more vigour than the bulk genome[1,2]. Several new, exciting examples have come into focus in the recent past which demonstrate that the DNA repair process is coupled with replication or transcription. The transcription factors or specially named coupling factors are shown to be implicated in each example[3,4]. Following the principle of economy, it seems plausible that the cell would use the protein factors used in one of the processes to fulfil the similar or common requirements of more than one process. For example, DNA scanning, melting and unwinding of the local regions underlying these processes can be performed by the same protein molecules in all the instances, which can later recruit or complex with different process-specific factors to generate diverse functional complexes. The cells follow exactly the same strategy in practice. As and when possible, the cell tries to use the available set of proteins/factors used in one of the reaction mechanisms, already present as a functional complex on the DNA rather than designing and producing new, separate components for another mechanism. These common factors may serve as connectors of the diverse processes to switch from one mechanism to the other and search for such factors has been on for a long time. Identification and studies on the properties of such a factor along with the cloning of its gene from the bacterial cell[4] have helped in the discovery of putative coupling factors from other eukaryotic cells as well. Research on complex reaction mechanisms of eukaryotic cells thus continues to get guiding leads from the relative simplicity of the mechanisms used by a prokaryotic cell.

Bacterial coupling factor

The nucleotide excision repair (NER) mechanism for DNA repair recognizes a wide range of damages and eliminates a broad spectrum of lesions thus serving as the major pathway of DNA repair. NER mechanism seems to be highly conserved among various species[5,6], wherein several sequential activities like the damage recognition, nucleolytic incision on both sides of the lesion, excision of the oligomer, repair synthesis and ligation of the new ends, by several gene products, all result in repair of the damage. While the details of the pathway are resolved in the microbe *E. coli*, with the identification of several *Uvr* gene products, the components of the pathway in eukaryotic cells remain elusive, exhibiting a more complex pattern due to the involvement of numerous genes (Table 1). The mfd (*m*utation *f*requency *d*ecline) gene product of *E. coli*, which was shown to be necessary for strand specific repair was found to do two things at once. It could displace a stalled RNA polymerase molecule at a lesion in an ATP-dependent manner, as well as bind to the damage recognition subunit of *UvrA* of the excision nuclease; and the repair of the transcribed strand could occur only during active transcription process. The mfd protein was also found to have helicase motifs and regions of sequence similarity to UvrB and RecG proteins. Thus the mfd gene product could fit the bill as the long sought after 'molecular matchmaker'[7] of transcription and repair and thus be rightly termed as the 'transcription-repair coupling factor' (TRCF)[4] of *E. coli*. It is not difficult to visualize that a similar coupling factor may also be present in eukaryotic cells. A gene product can be considered as a putative coupling factor if its requirement is shown for both transcription and repair. It may also have an ATP-dependent helicase activity and it may be capable of forming protein–protein bridges with its counterpart.

Eukaryotic TRCF

Clues from human genetic disorders

The wealth of information collected during the past two years gives a fresh insight into the old lessons from bacteria; also several phenomena appear as rediscoveries in a new context. This explosion in information has been possible through several detailed molecular genetic studies of the biochemical processes in yeast, Drosophila, rat and human genes/mutants. Details of the phenotypic and genetic analyses of some of the rare human diseases as well as analogous Drosophila and yeast RAD mutants have given deeper insights into the connection between DNA repair and transcription processes. For example, the RAD1/RAD10 complex of yeast (single strand nuclease) is involved in mitotic recombination stimulated by transcription[8], thus suggesting that transcription/repair complexes may be permanently associated. In humans, the inability to repair damaged DNA is associated with diseases, at least two of which have been studied by now. Both the diseases, namely Xeroderma pigmentosum (XP) and Cockayne's syndrome (CS) show phenotypes of wide varieties, unrelated to each other, due to defects lying in unidentified early steps of the pathway preceding incision and repair

Table 1. DNA repair genes of yeast and homologous genes in human repair deficiency diseases

Yeast NER genes	Major activity: established/suggested for		Human gene/mutant	Disease complementation group
	Yeast	Human homolog		
RAD14	Damage recognition, DNA binding, zinc finger motif	Zn finger, binds ss DNA UV damaged DS DNA	XPAC	XP-A
RAD25 (SSL2)	Essential, required for pol II transcription, helicase motifs	Helicase motifs, part of TFIIH, chromatin binding	XPBC/ERCC 3	XP-B
RAD4	Chromatin, DNA binding?		XPCC?	XP-C
RAD3	Essential, 5′-3′ helicase?, RNA–DNA helicase, subunit of factor b (TFIIH)	Essential helicase, DNA binding	XPDC/ERCC 2	XP-D
–	–	–	XPEC	XP-E
–	–	–	XPFC	XP-F
RAD2	SS specific endonuclease transcription inducible by UV	Functions in complexation with ERCC 4, ERCC 11, XPFC	XPGC/ERCC 5	XP-G
–	–	–	XPBC/ERCC 3 & XPDC/ERCC 2	XP/CS
RAD16 (SNF2)	Helicase?, DNA binding, Zn finger homology to RAD5 & 54	Helicase, NTP binding sites, non-essential	CSAC CSBC/ERCC 6	CS-A CS-B
Other yeast repair genes				
RAD10 also RAD1, RAD5, RAD7, RAD23 & 24	RAD 1/10 complex	C-terminus homologous to UvrA and UvrC, involved in recombination?	ERCC 1	

synthesis. Many of the human NER genes have been cloned by complementing the exicision repair deficiency of rodent cell lines and subsequently named as ERCC (*E*xcision *r*epair *c*ross *c*omplementing) genes. Some other genes of the pathway were, however, cloned by directly complementing the NER deficiency of XP cell lines. Several complementation groups of both the disorders are known (Table 1). The mutant phenotypes range from UV sensitivity in all the cases with (XP) or without (CS) skin cancer, neurological abnormalities, hypogonadism and retarded growth. Obviously, all these phenotypes cannot be attributed directly to DNA repair deficiency alone. Some unknown connections of the genes regulating these phenotypes with the repair deficiency seem to be the best explanations.

Studies during the last few months have successfully unravelled some of these connections. The question is whether the observed heterogeneity in the phenotypes could be due to some general connection of repair and transcription resulting in pleiotropic effects; if yes, what is the mechanism of this link. The task was difficult not only because the link between transcription and repair in eukaryotes is not known, but more so because the mechanism and intermediate steps of even the NER process itself still need to be worked out. In a major advancement, the gene of one of the repair proteins ERCC3, has been shown to code for one of the subunits of an essential basal transcription factor TFIIH in human cells[9]. TFIIH is one of the basal factors required by the enzyme RNA polymerase II to form a functional initiation complex. Several enzymatic activities such as ATPase, helicase and even a kinase activity were shown associated with this multi-subunit factor. TFIIH is also required for the formation of open initiation complex and for switching the complex from initiation to elongation mode[10]. Thus it will be interesting to see if ERCC3 is also found in the elongation complex. The same protein can couple repair and transcription during both the steps, or in other words the coupling of repair and transcription complex components may be processive in nature.

ERCC3 gene product in humans (same as the 89 kDa subunit of the basal transcription factor TFIIH[9]) is presumably a repair helicase involved in connecting the defects in a very rare combined form of Xeroderma pigmentosum complementation group B (XPBC) and Cockayne's syndrome. Phenotypes of the two rare human disease (XP and CS) are caused by partial loss of function due to mutant proteins. The XP allele of mutant ERCC3 gene makes a truncated protein. In the case of *E. coli*, mfd$^-$ is null phenotype in contrast to yeast where null allele is lethal. The truncated XPBC protein cannot work in NER (resulting in XP) with a simultaneous reduction in its capacity to function in concert with RNA polymerase II transcription machinery (resulting in CS). Thus a deficiency in transcription process itself rather than a defect in preferential repair of transcribed strand may be the cause of Cockayne's syndrome. The Drosophila homolog of ERCC3, *haywire* (66% identify) was shown to have many recessive lethal alleles like in XP. Viable alleles caused UV sensitivity and marginal expression displayed motor defects and reduced life span[11]. Using the analogy, a defect in transcription could also account for

various phenotypes of the *hay*^{nc2} mutant of Drosophila. Similarly, one of the NER genes of yeast, RAD25 (SSL2) having 54% sequence identity with ERCC3 was also recently shown to be required for transcription by RNA polymerase II[12]. The properties of the ERCC3 protein and presence of its homologs in various organisms suggest that despite comments[13] and counter-comments[14], ERCC3 may be considered the putative human TRCF by analogy.

XP arises largely due to autosomal, recessive lethal mutations causing extreme UV sensitivity and susceptibility to skin cancer. CS in contrast shows no cutaneous abnormality, retarded growth and severe progressive neurological degeneration[15]. One of the very rare disorders called Brittle hair trichothiodystrophy (TTD1BR) in XP-D was found to be a repair deficiency without photosensitivity. In the other extreme case of CS, cells were found to be defective in preferential repair of lesions on the transcribed strand of active genes, which permits quick resumption of transcription after UV exposure in normal conditions. Similarly, CS fibroblasts did not show any capacity to preferentially repair the active genes but were proficient in overall genome repair. Thus, in spite of normal range repair function, they were unable to recover their RNA synthesis after UV irradiation[6]. The mutation of the gene affected in CS-B patients, ERCC6 showed defect in preferential repair of active genes and the protein involved was found to have seven motifs characteristic of several proteins of the family of helicases[16]. ERCC6 is also proposed to be a human TRCF.

Clues from yeast mutants

Thus, several recent results suggest that there may be precise controls to carry out transcription and repair in concert. More direct results come from studies on yeast. It was reported sometime ago that the transcribed strand of the GAL7 gene is repaired 2–3 times faster than the non-transcribed strand or the overall genome[17]. The transcription complex thus seems to direct the repair machinery to the transcribed strand of active genes. This strand selectivity was lost in the temperature sensitive mutant of the largest subunit of yeast RNA polymerase II, rpb1-1, which was found to resemble the temperature sensitive rad25 mutant. Upon shifting to non-permissive temperature, it was seen that pol III transcription is not affected, while pol I transcription is hampered and pol II transcription undergoes rapid decline[12]. Effect on pol I transcription in rpb1-1 mutant could be explained due to the stringent-response. It is possible that the rad25 mutation produces inhibition of transcription of genes involved in the translation initiation process, resulting in the observed effects. RAD25, which was later found to be required for pol II transcription, contains the conserved ATPase/DNA helicase sequence motifs and the domains responsible for transcription and repair activity could be demonstrated to be separable[12]. In the beginning of this year, a report in *Nature* showed that another yeast NER gene RAD3 is essential for pol II transcription. The rad3-ts$_{14}$ mutant was also found to resemble rpb1-1 mutant and the gene product was shown to have DNA-dependent ATPase and helicase activities[18]. The latest piece of information has appeared in a recent issue of *Nature*[19] where the authors show that RAD25 associates with, and RAD3 purifies as, the subunit of yeast basal transcription factor b (homolog of human TFIIH) and that SSL2 is required for NER. This gives more direct proof that SSL2 (RAD25) binds to factor b of yeast through the other repair factor RAD3 which is actually a subunit of factor b and even has RNA-DNA helicase activity. This way, RAD25 can be accepted as the true TRCF of yeast. Mutations in human homolog of RAD3, ERCC2, result in XP-D, CS and trichothiodystrophy (Table 1); thus ERCC2 may also be a putative coupling factor. While none of suggested human TRCFs showed any strong sequence homologies with mfd of *E. coli*, it seems there may be more than one TRCF and more than one NER pathways in human cells, each having its own factor of coupling with transcription. Multiple pathways may also explain multiple complementation groups for apparently similar disorders.

While three clear cut models XP, CS and PIBIDS (TTDIBR) of repair defects in humans are available, *Drosophila haywire* can prove to be a more convenient model for these human diseases and along with yeast mutants, it may give valuable information on the basic mechanisms underlying transcription-repair coupling. Analysis of the function of a normal protein in a complementation assay using the mutant extract, or vice versa, can even help in deciphering the whole NER pathway[20–22], for which most of the proteins involved would be helicases or the nucleases. Several of the NER genes have been cloned by now. Thus it may soon be possible to replace a TFIIH component by one of the ERCC3 mutant allele product. It may also help to find out whether TFIIH activity is affected during elongation or initiation of transcription, which in turn will tell if TFIIH activity has any role to play in elongation. With the demonstration of the coupling of the transcription with repair it would also be important to study if all the proteins of transcription and repair coupling complex remain continually bound together as a complex, or whether the required proteins go on and off the complex as and when required under the signals from as yet unidentified additional component (or components). The cloning of rodent NER genes may make possible generation of repair defective transgenic mice to serve as better models for human diseases and study several phenomena related to ageing, neuronal degeneration, measurement of toxicity of a mutagen etc. It may also help us understand how the molecular defects may get translated into clinical phenotypes so diverse and heterogeneous, such as those found in the human disorders like repair deficiencies and how mutations in a single pathway may predispose cells to onset of cancer in some and not all cases.

These new and interesting results showing the link between repair and transcription give rise to a new class of disorders, according to which many of repair deficiency diseases can be classified as 'transcription syndromes'. This concept of transcription syndrome in humans is certainly striking and it may also create a new chapter in human genetics. It not only shows that extremely rare disorders can still be very revealing but also suggests that such disorders are necessarily rare, because they affect the very fundamental, basal activity of the cell. Thus it can be speculated that all extremely rare genetic disorders must be due to a defect at a crucial point in a very basic metabolic activity of the cell. And above all, those of us working on the basic process of transcription may find here a strong case to present as another example of the importance and relevance of basic research, when questioned by the staunch

supporters of the much demanded applied research!

1. Mellon, I. and Hanawalt, P. C., *Nature*, 1989, **342**, 95–98.
2 Mellon, I., Spivak, G. and Hanawalt, P. C., *Cell*, 1987, **51**, 241–249.
3. Xiong, Y., Zhang, H. and Beach, D , *Cell*, 1992, **71**, 505–514.
4. Selby, C. P. and Sancar, A , *Science*, 1993, **260**, 53–58
5. Hoeijmakers, J. H. J., *Trends Genet*, 1993, **9**, 173–177.
6. Hoeijmakers, J. H. J., *Trends Genet*, 1993, **9**, 211–217.
7. Sancar, A. and Hearst, J. E., *Science*, 1993, **259**, 1415–1420.
8. Sung, P., Reynolds, P., Prakash, L. and Prakash, S., *J. Biol Chem.*, 1993, **268**, 26391–26399.
9 Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H. J., Chambon, P. and Egly, J.-M., *Science*, 1993, **260**, 58–63
10. Lu, H., Zawel, L., Fisher, L , Egly, J -M. and Reinberg, D. *Nature*, 1992, **358**, 641–645.
11. Mounkes, L. C., Jones, R. S., Liang, B.-C., Gelbart, W. and Fuller, M. T., *Cell*, 1992, **71**, 925–937.
12. Qiu, H., Park, E., Prakash, L. and Prakash, S , *Genes Develop.*, 1993, **7**, 2161–2171.
13. Buratowski, S., *Science*, 1993, **260**, 37–38.
14. Sweder, K. S. and Hanawalt, P. C., *Science*, 1993, **262**, 439–440.
15. Friedberg, E. C., *Cell*, 1992, **71**, 887–889.
16. Troelstra, C., Van Gool, A., de Wit J., Vermeulen, W., Bootsma, D. and Hoeijmakers, J. H. J., *Cell*, 1992, **71**, 939–953.
17. Leadon, S. A and Lawrence, D. A., *J. Biol. Chem* , 1992, **267**, 23175–23182.
18. Guzder, S. N , Qiu, H , Sommers, C. H., Sung, P., Prakash, L. and Prakash, S., *Nature*, 1994, **367**, 91–94.
19. Wang, Z., Svejstrup, J. Q., Feaver, W. J., Wu, X , Kornberg, R. D. and Friedberg, E. C., *Nature*, 1994, **368**, 74–75.
20. Scherly, D., Nouspikel, T., Corlet, J., Ucla, C., Bairoch, A. and Clarkson, S G., *Nature*, 1993, **363**, 182–188.
21. O'Donovan, A. and Wood, R. D., *Nature*, 1993, **363**, 185–188
22. Habraken, Y., Sung, P., Prakash, L. and Prakash, S., *Nature*, 1993, **366**, 365–368

ACKNOWLEDGEMENTS. I wish to thank Prof. D. Balasubramanian for critically reading the manuscript and making useful suggestions.

Purnima Bhargava is in the Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India