Research News

The RNA world meets behavior: Afi I pre-mRNA editing in animals

Robert A. Reenan

Speculations on the genetic component of animal behavior have been fueled primarily by single-gene mutations that affect specific behaviors in model organisms. Pre-mRNA editing by adenosine deaminases acting on RNA (ADARs) provides an additional mechanism for introducing protein diversity and has primarily been observed in signaling components of the nervous system. Two recent reports of mutant mice and *Drosophila* deficient in ADAR activities provide further evidence that pre-mRNA editing has an ancient and primary role in the evolution of nervous system function and behavior.

The current avalanche of genomic sequence data has spurred *in silico* efforts to interpret it, mainly based on the comforting dogma that a given nucleotide triplet in a coding sequence will always encode a given amino acid in the final protein. However, for nearly a decade, adenosine-to-inosine (Afi I) pre-mRNA editing, a process which enzymatically modifies Afi I at single nucleotide positions in specific messages, has lurked in the background. 'Lurked' because the ribosome interprets I as G, posing an obvious threat to the predictive power of the genetic code[1]. The enzymes responsible for Afi I conversion, the ADARs, target only a few nucleotides per transcript molecule to change protein-coding potential (reviewed in Ref. 2). Importantly, only a handful of transcripts are known to be targeted by ADARs, and the full magnitude of Afi I editing of the transcriptome remains unknown. Although we understand the chemical basis and are learning about both the mechanism and the preferred targets for this type of editing, an essential question remains: why do organisms recode mRNAs enzymatically rather than simply incorporating those changes into the genome through mutation or the creation of alternative exons? What is the benefit of doing it this way?

Mode of discovery and function of the known targets of RNA editing

RNA editing was initially discovered in organelles and is now well documented (reviewed in Ref. 3). Nuclear pre-mRNA targets of Afi I editing have all been discovered serendipitously. Because inosine has base-pairing properties like those of guanosine, evidence for editing comes from Afi G differences among cDNA products or between cDNA and genomic sequences during the course of cloning. The first example of Afi I editing in an mRNA was found in the mammalian brain, in transcripts of the gene encoding the ionotropic glutamate receptor subunit, GluR-B. Other examples have appeared in numerous signaling components of the nervous systems of vertebrates and invertebrates, as well as in the genes encoding ADARs themselves (Fig. 1). Although some of these editing sites, like the Q/R and R/G sites of certain GluRs, are shared between paralogous genes, most are unique to the genes in which they reside.

Surprisingly, the few reports of specific Afi I pre-mRNA editing are all from nervous system targets, despite the fact that ADAR activities are detected in numerous non-nervous tissues. Also, inosine is detectable in polyA$^+$-containing mRNA in most tissues, with the highest levels found in brain messages[4]. Thus, there are probably many more undiscovered mRNA ADAR targets.

Where the consequences of RNA editing are known for given target genes, the changes in protein function are often

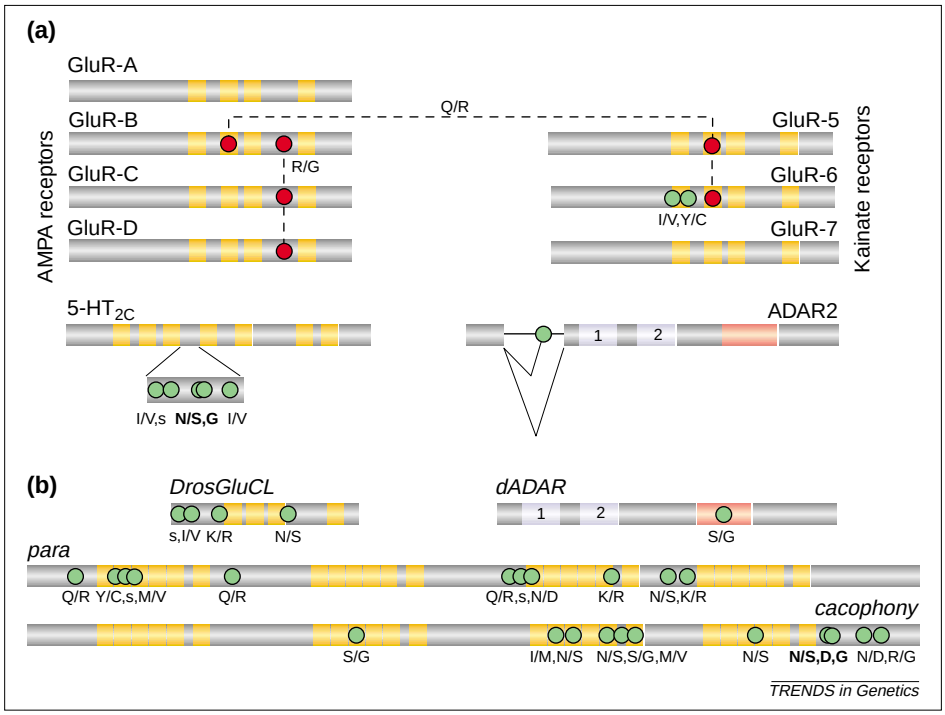

Fig. 1. Sites of Afi I pre-mRNA editing in mammals (a) and *Drosophila* (b). Circles represent editing sites that are at unique coding positions (green) or are in identical coding positions between paralogous genes (red). Note that one editing site (Q/R) is shared between more distant relatives of the glutamate receptor (GluR) families: the AMPA receptor GluR-B and the kainate receptors GluR-5 and GluR-6. The ADAR2 editing site is within noncoding sequences and influences alternative splicing. Paralogous sites are connected by dashed lines. Yellow boxes, the positions of transmembrane motifs relative to RNA editing sites; orange boxes, the catalytic domains of ADARs; numbers, double-stranded RNA-binding motifs in the ADARs. The effect of editing on coding potential is given below the particular editing site as 'unedited/edited'. Lower case 's' indicates that the editing introduces a silent change (for mammalian editing sites, see Ref. 23; for *Drosophila* editing sites, see Ref. 24). The *para* and *cac* genes encode *Drosophila* voltage-gated Na^+ and Ca^{2+} channels, respectively.

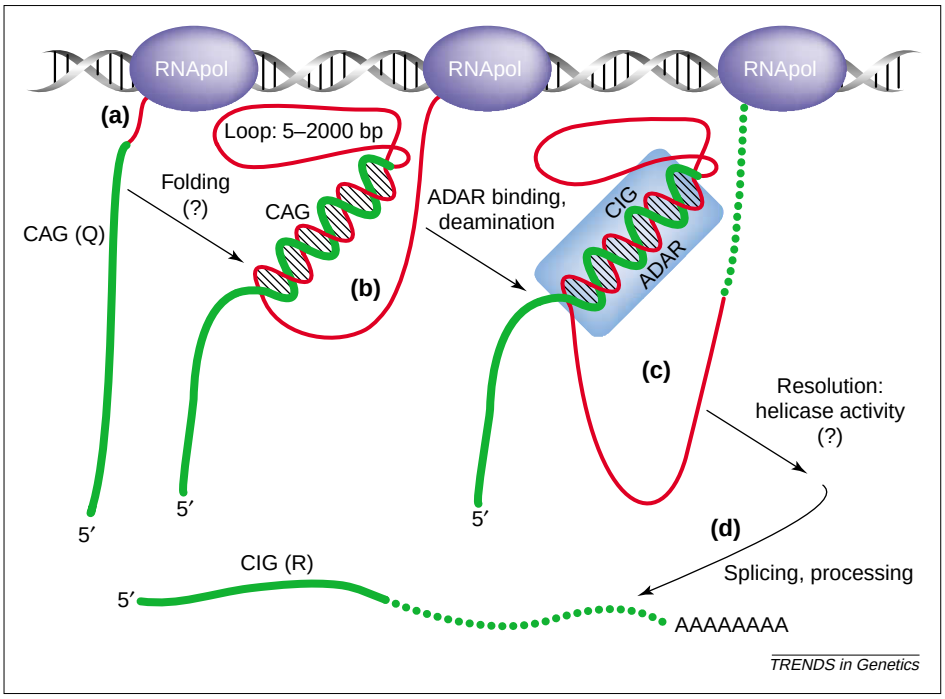

Fig. 2. The mechanism of Afi I editing for pre-mRNAs. RNA polymerase is shown in various stages of the transcription of an ADAR target gene. (a) A coding segment is transcribed into pre-mRNA and contains an adenosine destined to be modified in a glutamine codon (CAG). (b) Further transcription unveils the intronic editing site complementary sequence (ECS), which somehow base-pairs and forms a helix with the coding region, looping out intervening sequences. Exonic sequences are in green, and intronic sequences are in red. In some examples, the 5¢ end of the pre-mRNA might be tens of thousands of nucleotides long and contain many more exons and introns than are shown. (c) An ADAR binds to the duplex and deaminates Afi I recoding the message (Qfi R in this instance). (d) Further processing, which might involve helicase activity, to resolve the secondary structure and removal of spliceosomal introns.

profound. For instance, the editing of the GluR-B Q/R site is a major determinant of the Ca^{2+} permeability of multimeric GluR channels that incorporate the GluR-B subunit[5], whereas the editing of the GluR-B, -C and -D R/G sites affects the rate of recovery from desensitization of GluR channels[6]. Also, the efficiency of G-protein coupling of the serotonin receptor 5-HT_{2C} is greatly reduced in certain edited forms[7,8]. Editing can even regulate splicing. In the case of the gene encoding ADAR2 in mammals, editing by ADAR2 of its own pre-mRNA creates a splice acceptor site, leading to a splice form that is poorly expressed and introduces a different starting methionine into the ADAR2 open reading frame[9]. The *Drosophila* ADAR (dADAR) also appears to edit its own transcript, altering a specific amino acid position[10].

What selects an mRNA for editing?

The *in vitro* substrate for an ADAR is a double-stranded RNA of more than ~20 bp (Refs 11,12). The specificity of ADAR activity is dependent upon the extent of duplex complementarity. Perfectly matched duplexes are modified at more adenosine positions than duplexes that are interrupted by bulges or loops[13]. Consequently, it is thought that a major determinant of ADAR specificity is the local RNA structure around a targeted adenosine and that duplex imperfections position the ADAR.

But how are ADAR substrates formed in nascent pre-mRNAs *in vivo*? The formation of such a duplex was demonstrated in the GluR-B primary transcript at the Q/R editing site[14]. Base-pairing was demonstrated between coding sequences in the region of the modified adenosine and the editing site complementary sequence (ECS) residing in noncoding intron sequences (Fig. 2). Somehow, the bases of the ECS manage to pair and form a topologically constrained helix with exonic sequences in the nascent transcript, thus forming the ADAR substrate *in vivo*. This interaction can occur over large distances as some ECSs have been identified as far as 2000 bp from their associated editing site[15]. One implication of the location of ECS sequences is that RNA editing must occur before splicing. In addition, it has been proposed that helicase activity is necessary for the resolution of these structures for proper splicing to occur[16].

Why bother with RNA editing?

Why not just change a nucleotide through mutation? One intrinsic advantage of editing a nucleotide over having change 'hard-wired' into the genome through mutation is the regulation of the degree to which a coding position is modified within messages. Certain pre-mRNA editing sites vary greatly in the frequency with which their editing is detected *in vivo*, ranging from a few percent to nearly 100%. Thus, editing introduces levels of expression intermediate to the usual genetic variation (i.e. 0, 1 or 2 copies), possibly conferring selective advantage. In addition, because the few known ADAR target mRNAs are commonly edited at multiple positions independently (Fig. 1), the combinatorial effect of editing greatly increases the number of protein products that can be generated from an edited gene.

Another difference between editing and normal genetic variation is the potential for spatial and temporal regulation. Because an ADAR enzyme introduces modifications into mRNAs, editing is thus dependent on the amount and location of that protein. Indeed, developmental regulation of pre-mRNA editing occurs in organisms as diverse as fruit fly and rat[6,17,18]. In both cases and at numerous sites, editing of transcripts increases during the course of development. Furthermore, in the mammalian brain, RNA editing of GluRs and 5HT_{2C} is subregion specific[7,19,20]. Also, ADARs themselves have different tissue expression profiles[21].

The utilitarian aspects of RNA editing aside, what is the significance of Afi I editing for organisms? Perhaps the most studied example of Afi I RNA editing is the mammalian GluR-B Q/R site, which is edited at nearly 100%. Importantly, mice heterozygous for an allele of GluR-B that is uneditable at the Q/R site display postnatal neurological defects and early death[22]. This is the only information available on the significance of modifying an editing site *in vivo*. However, two recent studies tackle the issue of the global significance of RNA editing through the use of mutant animals that lack ADAR activities.

Higuchi *et al*. showed that mice deficient in ADAR2 display phenotypes very similar to animals mutant only for GluR-B Q/R site editing[23]. At the molecular level, ADAR2-mutant mice show decreased editing at many sites in transcripts of several GluR subunit genes and 5-HT_{2C}, as well as ADAR2 itself.

There is, however, continued editing of some substrates in the ADAR2-mutant mice indicating that ADAR1, ADAR3 or unidentified ADARs also have editing roles *in vivo*. This study also showed a significant reduction in the ability of GluR-B transcripts to undergo proper splicing at the Q/R site. Surprisingly, most phenotypes of the ADAR2-knockout mice are rescued by homozygosity for a version of the GluR-B gene that is pre-edited at only the Q/R site. Higuchi *et al*. state that future studies will be aimed at determining the phenotypic consequences of under-editing other ADAR2 targets in the rescued mice, implying that there might be more subtle behavioral deficits.

In the second study, Palladino *et al*. describe *Drosophila* mutants completely lacking ADAR activity[24]. Surprisingly, null mutants of *dADAR*, the only ADAR homolog in *Drosophila*, are morphologically normal, displaying a range of adult behavioral defects and phenotypes including motor deficits, mating defects, obsessive cleaning, flightlessness, temperature-sensitive paralysis and age-dependent neurodegeneration. These phenotypes are consistent with perturbations in the known targets of dADAR – voltage- and ligand-gated ion channels of the nervous system (Fig. 1), all of which are completely unedited in *dADAR* mutants. Unlike mouse ADAR2 knockouts, however, *dADAR*-mutant flies are not short-lived under ideal conditions.

Because most known targets for specific Afi I RNA editing appear to be signaling components of vertebrate and invertebrate nervous systems, and the phenotypes of animals lacking ADAR activities are primarily neurological or behavioral, RNA editing has obviously evolved a position of high importance in modifying the function of the nervous system. The number of known ADAR pre-mRNA editing sites and the potential for many more implies that any one particular editing site will have subtle effects on nervous system function and behavior. So, how can we explain the predominating effects of the GluR-B Q/R editing site in mice? Evolutionary studies provide a clue.

Evolution of the targets

Taking the GluR genes of mammals as an example, many GluR genes have a different set of editing sites (Fig. 1). In addition, the frequency of editing at each site is characteristic for each gene. Only the GluR-B gene is edited completely at the Q/R site, whereas the transcripts encoding GluR-A,-C,-D and GluR-7 are not edited at the Q codon of the Q/R site at all. Even more intriguing are the differences between the Q/R editing sites of AMPA receptors (GluR-B) and kainate receptors (GluR-5, -6). The ECS for GluR-B Q/R is 300 bp downstream of the editing site, whereas for both GluR-5 and GluR-6, the ECS is more than 1800 bp downstream[15]. Moreover, both the intron–exon boundaries and the predicted local duplex secondary structures are very different for these two groups. It is tantalizing to consider that editing of the Q/R site of the GluR-B and GluR-5, -6 paralogs might be an example of convergent evolution where editing of the same amino acid arose independently through the genesis of two different RNA structures. This idea is strengthened by recent analyses in invertebrates and vertebrates demonstrating that when RNA editing is conserved in orthologous genes, both the intron locations and the predicted local duplex structures necessary for RNA editing are very highly conserved[18,25].

Why might convergent evolution of strictly paralogous editing sites have such different physiological consequences? Given the evidence for conservation of editing-site RNA structures, a newly evolved editing site would arise by chance when unconstrained intronic sequences changed enough to form a partial duplex surrounding the edited nucleotide. This initial duplex would probably be an inefficient ADAR substrate. However, selectively advantageous editing sites would be improved over time by additional changes to noncoding sequences that increased the size of the RNA duplex and introduced bulges to position the ADAR correctly. Such changes would optimize the pre-mRNA as a specific ADAR substrate and would also be reflected in the stability of the RNA substrate over evolutionary time.

Because most sites are edited at much less than 100%, what explains the unusually high level of editing at the Q/R site of GluR-B genes in mammals? Undoubtedly, an early GluR-B gene was edited at an optimal, yet intermediate, level at the Q/R site. It is possible that mutations in the gene encoding GluR-B then occurred that allowed higher levels of editing of the GluR-B Q/R site. Such changes could affect the function of channels incorporating GluR-B subunits, or perhaps the post-transcriptional processing of GluR-B transcripts, because editing at the Q/R site seems necessary for proper processing[23]. Through a 'ratcheting' mechanism in which successive neutral or slightly deleterious changes allowed higher compensatory levels of editing at the Q/R site, GluR-B editing achieved high levels and high physiological significance. Because intermediate levels of GluR-B Q/R editing are highly deleterious and are the principal defect in ADAR2-knockout mice, such a model is attractive. Further evidence on the prevalence and level of editing of GluR-B orthologs in other vertebrates would help to answer this question, especially if GluR-B Q/R sites with intermediate levels of editing were found. Interestingly, two fish GluR-B orthologs encode R at the position of the GluR-B Q/R editing site in mammals[26]. Does this represent the ancestral state, or were high levels of editing of a fish ancestor 'hard-wired' into the genome through missense mutation?

Outlook

The recent results reporting the effects of ADAR deficiency in animals highlight the global significance of ADAR-mediated RNA editing and indicate several areas for future research. First, determining the precise effects of any particular RNA editing site will be of interest and might be aided by a combined evolutionary approach. Second, identifying potential ADAR targets will certainly require more detail about what ADAR enzymes recognize in their substrates. Hopefully, new bioinformatic tools can then be applied to the increasing amounts of genomic sequence data to identify potential new ADAR substrates. Third, the implications of ADAR-mediated editing for genetic disease are particularly intriguing. Not only are ADARs themselves a potential target for inactivating or gain-of-function mutations that might have global effects, but mutations in noncoding regions involved in editing of ADAR target genes can also be envisaged. Such mutations might have subtle but physiologically significant effects that are important in inherited neurological disorders. A full accounting of the effects of post-transcriptional processing, especially Afi I RNA editing, in the post-genomic era clearly presents a significant challenge!

Note added in proof

The gene encoding ADAR1 has recently been shown to be an essential gene in mice[27]. Surprisingly, these studies also revealed that ADAR1 was haploinsufficient. Mice with one dose of ADAR1 died as embryos with no obvious abnormalities. However, further analysis suggests that ADAR1 is involved in embryonic erythropoiesis. Although the targets of ADAR1 in embryos are unknown, ADAR1 haploinsufficient teratomas composed mostly of nervous tissue appear to edit the known adult targets of ADAR1 at a reduced level. These data stand in stark contrast to the roles of ADAR2 in mammals and dADAR in *Drosophila*.

References

1 Basilio, C. *et al.* (1962) Synthetic polynucleotides and the amino acid code, V. *Proc. Natl. Acad. Sci. U. S. A.* 48, 613–616
2 Rueter, S.M. and Emeson, R.B. (1998) Adenosine–inosine conversion in mRNA. In *Modification and Editing of RNA* (Grosjean, H. and Benne, R., eds), pp. 343–361, ASM Press
3 Simpson, L. (1999) RNA editing – an evolutionary perspective. In *The RNA World* (2nd edn) (Gesteland, R.F. *et al.*, eds), pp. 585–608, CSHL Press
4 Paul, M.S. and Bass, B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* 17, 1120–1127
5 Sommer, B. *et al.* (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 67, 11–19
6 Lomeli, H. *et al.* (1994) Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* 266, 1709–1713
7 Burns, C.M. *et al.* (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 387, 303–308
8 Niswender, C.M. *et al.* (1999) RNA editing of the human serotonin 5-hydroxytryptamine 2C receptor silences constitutive activity. *J. Biol. Chem.* 274, 9472–9478
9 Rueter, S.M. *et al.* (1999) Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80
10 Palladino, M.J. *et al.* (2000) *dADAR*, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* 6, 1004–1018
11 Nishikura, K. *et al.* (1991) Substrate specificity of the dsRNA unwinding/modifying activity. *EMBO J.* 10, 3523–3532
12 Polson, A.G. and Bass, B.L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.* 13, 5701–5711
13 Lehmann, K.A. and Bass, B.L. (1999) The importance of internal loops within RNA substrates of ADAR1. *J. Mol. Biol.* 291, 1–13
14 Higuchi, M. *et al.* (1993) RNA editing of AMPA receptor subunit GluR-B: a base-paired intron–exon structure determines position and efficiency. *Cell* 75, 1361–1370
15 Herb, A. *et al.* (1996) Q/R site editing in kainate receptor GluR5 and GluR6 pre-mRNAs requires distant intronic sequences. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1875–1880
16 Reenan, R.A. *et al.* (2000) The *mlenapts* RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the *para* Na^+ channel transcript in a region of RNA editing. *Neuron* 25, 139–149
17 Bernard, A. and Khrestchatisky, M. (1994) Assessing the extent of RNA editing in the TMII regions of GluR5 and GluR6 kainate receptors during rat brain development. *J. Neurochem.* 62, 2057–2060
18 Hanrahan, C.J. *et al.* (2000) RNA editing of the *Drosophila para* Na^+ channel transcript: evolutionary conservation and developmental regulation. *Genetics* 155, 1149–1160
19 Melcher, T. *et al.* (1996) RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J. Biol. Chem.* 271, 31795–31798
20 Bernard, A. *et al.* (1999) Q/R editing of the rat GluR5 and GluR6 kainate receptors *in vivo* and *in vitro*: evidence for independent developmental, pathological and cellular regulation. *Eur. J. Neurosci.* 11, 604–616
21 Paupard, M.C. *et al.* (2000) Patterns of developmental expression of the RNA editing enzyme rADAR2. *Neuroscience* 95, 869–879
22 Brusa, R. *et al.* (1995) Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* 270, 1677–1680
23 Higuchi, M. *et al.* (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 406, 78–81
24 Palladino, M.J. *et al.* (2000) Afi I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. *Cell* 102, 437–439
25 Aruscavage, P.J. and Bass, B.L. (2000) A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. *RNA* 6, 257–269
26 Kung, S. *et al.* (1996) Characterization of two fish glutamate receptor cDNA molecules: absence of RNA editing at the Q/R site. *Mol. Brain Res.* 35, 119–130
27 Wang, Q. *et al.* (2000) Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* 290, 1765–1768

R.A. Reenan
Dept of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030, USA.
e-mail: rreenan@neuron.uchc.edu

Identification and analysis of eukaryotic promoters: recent computational approaches

Uwe Ohler and Heinrich Niemann

The DNA sequence of several higher eukaryotes is now complete, and we know the expression patterns of thousands of genes under a variety of conditions. This gives us the opportunity to identify and analyze the parts of a genome believed to be responsible for most transcription control – the promoters. This article gives a short overview of the state-of-the-art techniques for computational promoter localization and analysis, and comments on the most recent advances in the field.

Understanding gene regulation is one of the most exciting topics in molecular genetics. To learn how the interplay among thousands of genes leads to the existence of a complex eukaryotic organism is one of the great challenges. The quantity of information gained in the sequencing and gene expression projects both requires and enables us to use computers to solve this problem.

Promoter sequences are crucial in gene regulation. For the purposes of this paper, we define a promoter as the region proximal to the transcription-start site (TSS) of genes transcribed by RNA polymerase II; we exclude distal regions such as enhancers. Here we outline the recent developments in two areas of bioinformatics that deal with promoters: the general recognition of eukaryotic promoters, and the analysis of these regions to identify the regulatory elements in them. These analyses are the first step towards complex models of regulatory networks. We focus on the computational point of view and leave a more elaborate description, especially of the underlying biology, to the cited papers and reviews.

Analyzing promoters to find unknown regulatory elements

The interest in promoter analysis received a great boost with the arrival of microarray gene-expression data. Once you have a group of genes with a similar expression profile (e.g. those that are activated at the same time in the cell cycle[1]), a natural assumption is that this profile is, at least partly, caused by and

  PII: S0168-9525(00)02174-0