Journal of Bioenergetics and Biomembranes, Vol. 35, No. 1, February 2003 (© 2003)

Molecular Evolution of Nucleoside Diphosphate Kinase Genes: Conserved Core Structures and Multiple-Layered Regulatory Regions

Naoshi Ishikawa,[1,2] Nobuko Shimada,[1] Yohko Takagi,[1] Yasushi Ishijima,[1] Mitsugu Fukuda,[1] and Narimichi Kimura[1]

Received August 16, 2002; accepted October 4, 2002

Genomic data regarding the nucleoside diphosphate (NDP) kinase genes have been accumulated from diverged phyla. Comparison of their regulatory sequences have shed light on the multiple facets of gene regulation systems. Phylogenetic studies, including CpG island and intron-mapping, and homologous sequence comparison, have suggested that the regions of the major mammalian genes, the ortholog (rat α or *nm23-H2*) and its paralog (rat β or *nm23-H1*), have been constructed by a stepwise gain and loss of alien genes resulting in "multiple-layered" regulatory systems. They contain representative cis-elements for the constitutive, stage/lineage-specific, and early response expression. These elements' binding capacities to nuclear proteins were confirmed by electrophoretic mobility shift assay. Further, these regulatory systems generate heterogeneous mRNA at the 5′ untranslated region, which influences their own translation efficiencies. In terms of this process, the transcription system would control another layer of gene expression: posttranscriptional (translational) regulation.

KEY WORDS: Nucleoside diphosphate kinase (NDP kinase); nm23; *awd*; gene family; cis-elements; transcription factor; gene evolution; ortholog; paralog.

INTRODUCTION

More than a decade has passed since the discovery of the first nucleoside diphosphate (NDP) kinase genes, *nm23* in mouse (Steeg *et al.*, 1988), and *awd* in fruit fly *Drosophila melanogaster* (Biggs *et al.*, 1988; Dearolf *et al.*, 1988). Initially, the molecular properties of these gene products were unknown, but they drew attention for their intriguing biological properties as a tumor metastasis suppressor and an essential element of fly development, respectively. Soon, other groups' cloning data, one from peptide-based cloning of the rat gene (Kimura *et al.*, 1990), two other functional clonings of the slim mold gene (Lacombe *et al.*, 1990) and *Myxococcus* gene (Muñoz-Dorado *et al.*, 1990), revealed that all of these genes belong to an orthologous gene that encode an essential house keeping enzyme, NDP kinase. To date, in the human genome, eight genes have been identified as members of the NDP kinase gene family (reviewed in Lacombe *et al.*, 2000). Human NDP kinase genes were classified into two groups, Groups I and II, according to their genomic architecture and phosphotransferase activity (Lacombe *et al.*, 2000). All isoforms of the Group I (*nm23-H1*, *H2*, *H3*, and *H4)* possess well conserved NDP kinase active site motifs and are catalytically active, whereas the Group II genes (*nm23-H5*, *H6*, *H7*, and *H8*) provide highly divergent sequences from the Group I genes. The NDP kinase active site motifs. The NDP kinase active site motifs of Group II are not strictly conserved, and only *nm23-H6* has been reported to express NDP kinase activity (Tsuiki *et al.*, 1999). In mammalia, two major isoforms of NDP kinase have been reported from humans, mice, and rats (Gilles *et al.*, 1991; Shimada *et al.*, 1993; Stahl *et al.*, 1991; Urano *et al.*, 1992). Generally, more than 90% of cellular NDP kinase enzymatic activity is supplied by them. They are encoded by distinct genes arranged in a tandem array (Shimada *et al.*, 1993, see Fig. 1(a)). This configuration

[1] Cellular Signaling Research Group, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakaecho, Itabashi-ku, Tokyo 173-0015, Japan.

[2] To whom correspondence should be addressed; e-mail: naoshi@tmig.or.jp.

0145-479X/03/0200-0007/0

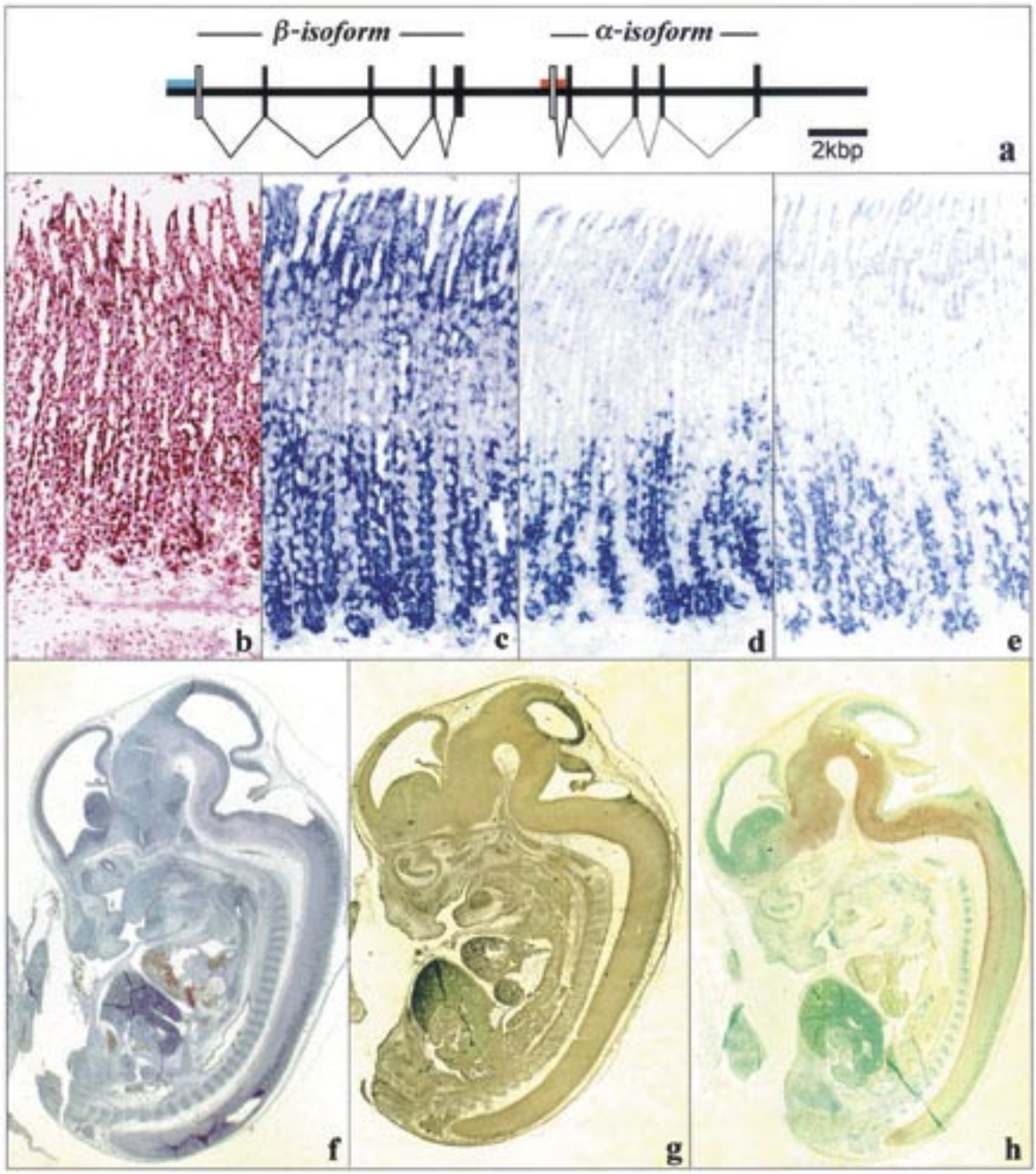

Fig. 1. Genomic structure of rat NDP kinase genes for two major isoforms, α and β, and the isoform specific expression of the mRNAs and the proteins. *Upper panel* – (a) The genomic segment containing two genes for the rat NDP kinase α and β isoforms are shown. The exons are depicted by boxes, open boxes represent untranslated exons and closed boxes represent coding exons, respectively. The transcription initiation windows for α and β genes are depicted by red and blue lines, respectively. *Middle panels* – (b–e) In situ hybridization analysis of rat gastric mucosa using isoform/group specific probes. Serial thin frozen sections of stomach (pars glandularis) derived from a 9-month-old Wistar rat were subjected to *in situ* hybridization. The hybridization was performed using three kinds of RNA probes labeled with digoxgenin, which were complementary to the group 1 type α untranslated region (c), the group 2 type α untranslated region (d), and β untranslated region (e). An adjacent section was stained by hematoxylin and eosin (b). *Lower panel* – (f–h) Immunohistostaining of 17-day *postcoitum* Wistar rat embryo by isoform specific monoclonal antibodies. The rat embryo was frozen in liquid nitrogen, and serial sections were stained by hematoxylin and eosin (f), and immunostained by α specific (g) and β specific (h) monoclonal antibodies. The sections were counter-stained with methylgreen.

demonstrated that these two genes were generated by a gene duplication event from an ancient ortholog, but precise analyses and the meaning of the gene expansion event have not been resolved (Shimada *et al.*, 1993). A couple of phylogenetic studies on NDP kinase gene family have been performed (Lacombe *et al.*, 2000; Troll *et al.*, 1993), but the evolutionary process to yield such a complex gene family largely remains to be determined.

In regard to gene expression, orthologous, or major isoforms of NDP kinase are expressed ubiquitously but in a spatiotemporal specific manner, and also are reportedly increased at specific life stages of several organisms and in various types of cells under certain circumstances; for example, during formation of the imaginal discs of *Drosophila melanogaster* (Biggs *et al.*, 1990), in murine systemic organs at their organogenesis stages (Lakso *et al.*, 1992), in concanavalin A-stimulated human T lymphocytes (Keim *et al.*, 1992), in human diploid fibroblasts when immortalized by various reagents (Ohneda *et al.*, 1994), and in the carcinogenic stages of certain human tumors. On the other hand, decreased expression of NDP kinase has been reported in the slime mold *Dictyostelium discoideum* when aggregation and development of the multicellular organization was triggered by starvation (Wallet *et al.*, 1990) and in the case of tumor metastases in experimental animal systems and in certain human tumors (De La Rosa *et al.*, 1995; Fukuda *et al.*, 1996; Leone *et al.*, 1991). Gene function *per se* resides in the protein encoded by the gene, whereas appropriately regulated gene expression is also essential to exert biological function. Indeed many reports have suggested that the increasing or decreasing expression of NDP kinase genes may cause the perturbation of various biological and/or pathological processes. To date, several cis-elements (or motifs) in 5′ upstream regions of the NDP kinase genes have been reported as putative regulatory elements for the gene expression, though the actual molecular interaction between such elements and transactivating factors largely remains to be elucidated.

This review focuses on NDP kinase gene structures from various organisms, particularly on their gene regulatory regions from an evolutionary viewpoint. Comparative study exploring the regulatory system of the genes could offer an intriguing perspective on the NDP kinase gene family regulation and function.

NDP KINASE GENE FAMILY: AFTER A LONG JOURNEY OF MOLECULAR EVOLUTION

Explosively accumulated genome data from various species are helping us to understand the essence of life. As a universal enzyme at a pivotal position of nucleotide synthesis pathway, NDP kinase has drawn special interest during the present genomic era. Highly conserved orthologous genes for NDP kinase have been isolated from every organism in three major domains: eubacteria, archaea, and eucarya. Recent phylogenetic studies have refined the bacterial domain phylogeny and have determined that the mycoplasma taxon, which is exceptionally lacking an NDP kinase orthologous gene, is located as a branch of the bacillus group that possesses an NDP kinase olthologous gene (reviewed in Wolf *et al.*, 2002). Consequently the last common ancestor might have a single ancestral NDP kinase gene (Fig. 2).

Phylogenetic study of the genes has revealed some intriguing features. First, unicellular organisms generally possess only one gene encoding NDP kinase, for example *ndk* in eubacteria, *E. coli* (Hama *et al.*, 1991), *#1265(ndk)* in methanogenic archaeon *Methanococcus jannaschii* (Bult *et al.*, 1996), *ynk* in yeast *Saccharomyces cerevisiae* (Fukuchi *et al.*, 1993), *ndk1* in fission yeast *Schizosaccharomyces pombe* (Izumiya and Yamamoto, 1995). No paralogous gene has been reported in these organisms. Second, most multicellular organisms possess multiple divergent paralogous genes beside an ortholog that provides most of the NDP kinase enzymatic activity. For example, slime mold *Dictyostelium discoideum*, known as the most primitive multicellular organism, has two genes, *ndkC* and *ndkM*, coding the cytosolic localized enzyme and mitochondoria-localized one, respectively (Troll *et al.*, 1993). The fruit fly has reportedly three, and the human genome bears at least eight related genes supposed to form a complex NDP kinase gene family. Third, major gene-divergent events most probably caused by gene duplications seem to have occurred at and after the explosive division of the three major domains: fungi, plants, and animals. The number and diversity of the gene family has seemingly been increased in association with the complexity of the body plan; therefore, to know how these paralogous genes arose and what new functions these genes perform is one of the major interests of this research field.

Obtaining a new gene copy is mainly achieved by two mechanisms: (1) gene duplication by unequal crossing-over or translocation of the genomic segment (2) polyploidization (Ohno, 1967). In the former case, the primary structure of the genomic copy might be essentially identical to the template partner, but under a certain constraint, conserved ortholog and diverged paralog are assumed to be generated. Interestingly, most of the paralogous NDP kinase genes possess two characteristic features compared with the orthologs as follows: (1) divergent 5′ and/or 3′ extension and/or deletion, in particular 5′ extension (for

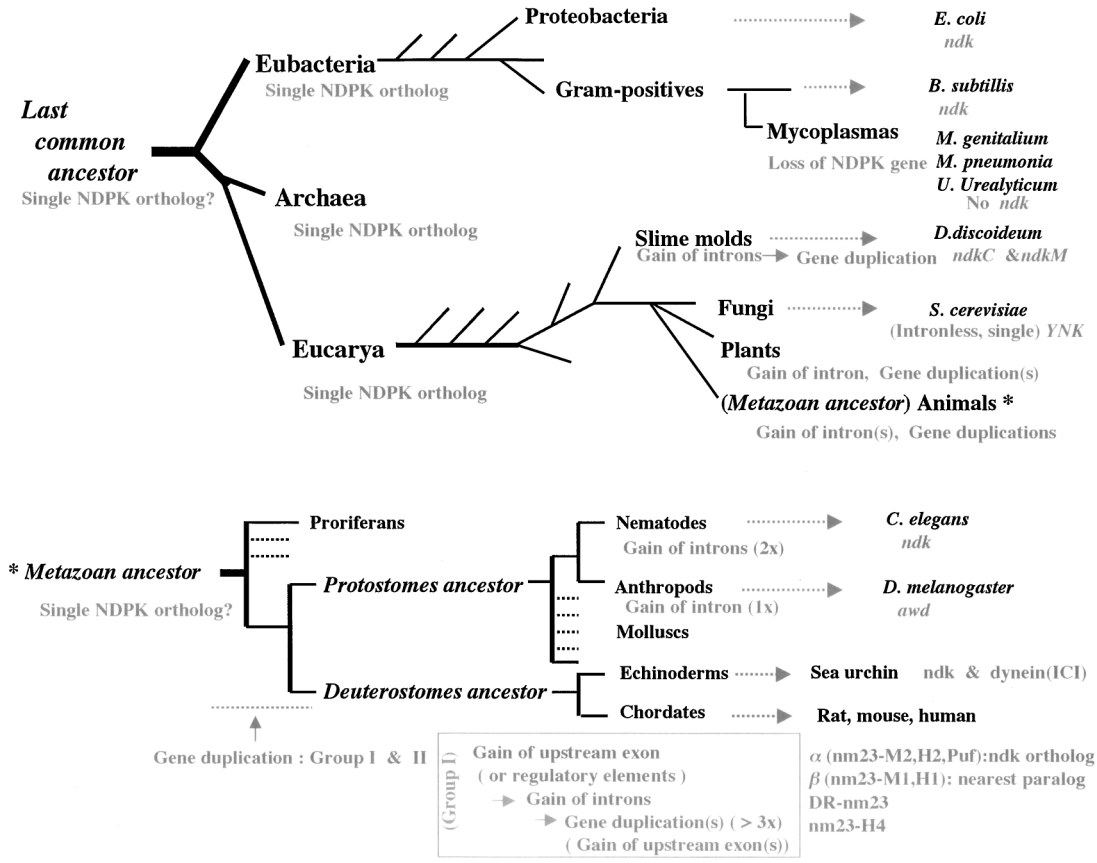

Fig. 2. Molecular evolution of NDP kinase gene family. A phylogenic tree was drawn following the previously hypothesized trees based on 18S ribosomal RNA (Adoutte *et al.*, 1999; Wolf *et al.*, 2002). Evolutional process of animal is closed up in the lower panel. Hypothetical events occurred in NDP kinase genes are denoted in gray. A gene duplication resulting in generation of Groups I and II of NDP kinase genes probably occurred before or around the metazoan radiation. A massive gene duplication might take place in the Deuterostomes lineage.

examples, *ndkM* in slime mold, and *DR-nm23/nm23-H3* in humans), (2) specified expression patterns (testis-specific expression of *nm23-H5* and *H7*, mitochondorial targeting of *ndkM* and *nm23-H4*). Since the main difference between the original gene and its copy is supposed to exist only for surrounding sequences at the initial stage, the paralogous gene might have an additional 5′-exon or regulatory region that possessed a certain common feature in order to compromise with the ortholog. The increased numbers of NDP kinase genes in amphibia and plants may be caused partly by the second mechanism. Another mechanism to create a new gene family is the integration of mRNA by reverse transcriptase resulting in a processed (intron-less) gene (see rhodopsin-type receptor family). Although numerous processed genomic segments that are closely related with the α or the β mRNAs were found in rat genome (Ishikawa *et al.*, 1992; Shimada *et al.*, 1993), no transcriptionally active gene has been identified to date.

NDP kinase genes universally reside among the three domains: eubacteria, archaea, eucarya; therefore, these genes were utilized as a good parameter for molecular evolution study (Doolittle *et al.*, 1996). A couple of phylogenetic studies on NDP kinase gene evolution *per se* have also been performed (Lacombe *et al.*, 2000; Troll *et al.*, 1993). However, recent phylogenetic studies have criticized the simple comparison, because of recurrent substitution resulting in underestimation of the distance (reviewed in Schopf, 1998), different substitution rate between ortholog and paralog (Wolf *et al.*, 2002), species-dependent substitution rates, and so forth. Further meticulous studies accounting for such factors are needed.

LOSS OF NDP KINASE: NATURAL OCCURRENCE AND ARTIFICIAL DISRUPTION

Mycoplasmas are members of the class of parasitic bacteria that lack a cell wall. One of the mycoplasmas,

M. genitalium, has the smallest known genome of any living organism. Thus, *M. genitalium* shares a special position for considering the minimal essential elements of life. Surprisingly this simplest genome doesn't contain NDP kinase (Fraser *et al.*, 1995). To date, a couple of the mycoplasmas' complete genomes have been reported, and it has been revealed that all of them lack the ortholog in their genomes (Glass *et al.*, 2000; Himmelreich *et al.*, 1996). The question arises as to whether NDP kinase is essential for life. Recent studies on bacterial phylogeny demonstrated that the mycoplasma clade was derived from a Bacillus group (low GC Gram positive) ancestor (Wolf *et al.*, 2002). A representative Bacillus *Bacillus subtillis* genome contains the orthologous gene *ndk* (Kunst *et al.*, 1997); therefore, lack of the NDP kinase orthologous gene in mycoplasmas was most probably due to the loss of the gene during parasitic adaptation of the ancestral bacteria. It should be noted that typical symbiotic organism, mitochondorion, has also lost NDP kinase. It is an intriguing question as to what constraining force worked upon the mycoplasma ancestor during parasitic adaptation.

Artificial disruption of a gene is a powerful tool to elucidate the gene's function. Trials for targeted disruption of the NDP kinase gene have been performed in several organisms, including *Myxococcus Xanthus* (Muñoz-Dorado *et al.*, 1990), *E. coli* (Hama *et al.*, 1991), the yeast *Saccharomyces cerevisiae* (Fukuchi *et al.*, 1993), and the fission yeast *Schizosaccharomyces pombe* (Izumiya and Yamamoto, 1995). It was reported that the *ndk* gene is essential for the Myxococcus cell growth. However, to our surprise, *E. coli*, the yeast, and the fission yeast whose sole NDP kinase gene was knocked-out have survived without remarkable phenomenological changes in their organisms within the extent examined. However, since afterwards the NDP kinase null-line of *E. coli* provides a mutation prone phenotype, NDP kinase has been reassigned as a novel mutator gene (Lu *et al.*, 1995). Since no species survives naturally without NDP kinase, this enzyme should be essential to the long-term survival for all organisms except for obligate parasitic organisms. Further study is needed to elucidate the relationship between NDP kinase and life. Arnaud-Dabernat *et al.* (2003) reports the result of gene targeting in a murine system in the present issue.

SPECIES-SPECIFIC LOCATION OF INTRONS

When and how the first intron was generated are still controversial subjects. The major arguments between the "introns-early" hypothesis (Gilbert, 1987) and the "introns-late" or transposon hypothesis (Cavalier-Smith, 1985) have been focused on the origin of introns (review in Trotman, 1998). Despite of these unsettled controversies, exon–intron structures in the homologous genes have been conserved in a certain manner. Therefore, the exon–intron structure of the gene would provide another hallmark of gene evolution. The intron numbers have generally increased as species evolution has proceeded. The insertion module of the intronic sequence in the representative NDP kinase genes is summarized as follows (Fig. 3): (1) unicellular organisms such as *E. coli* and *S. cerevisiae* possess one undivided NDP kinase gene (Fukuchi *et al.*, 1993; Hama *et al.*, 1991), (2) slime mold *D. discoideum* bears two genes coding for cytosolic and mitochondrial type enzymes, both of them having two intervening sequences at identical locations and additional two introns in the latter gene (Lacombe *et al.*, 1990; Troll *et al.*, 1993), (3) *D. melanogaster awd* has a single intron (Biggs *et al.*, 1988; Dearolf *et al.*, 1988; Adams *et al.*, 2000), (4) *C. elegans ndk* has two introns, the inserted positions of which are unique to this species, (5) the mammalian genes for two major isoforms (*nm23-H1* and *H2* in human, α and β in rat; *nm23-H1* corresponding to β and *H2* corresponding to α, respectively) contain four introns at completely conserved locations (Shimada *et al.*, 1993, see Fig. 1(a)).

Since the *awd* locus of the fruit fly has one intronic sequence exactly corresponding to the location of the second intron of the rat major genes, the *awd* gene is supposed to represent an archetype of the mammalian NDP kinase gene, and a duplication of the ancestral gene could have occurred in the chordate ancestor after the separation of these species (Ishikawa *et al.*, 1997; Shimada *et al.* 1993). On the other hand, the completed genome sequencing of nematode *C. elegans* revealed that the worm NDP kinase gene contains two introns, both of which are located in the species-specific positions (The *C. elegans* Sequencing Consortium, 1998, see Fig. 3). This fact raised questions regarding the previous simple notion. The study of species evolution, phyrogeny, is a major subject of biology, and many hypothetical "phylogenic trees" have been constructed. Traditional phylogenic studies have postulated that the ancestor of nematodes (then classified as Pseudocoelomates) would be separated from a coelomate's ancestor. However, recent studies based on 18S ribosomal RNA comparison have proposed that the common ancestor of the nematodes and anthropods (classified as Ecdysozoans' ancestor) would be separated from Protostomes' common ancestor after division of Protostomes' from Deuterostomes' ancestor (Fig. 2 is drawn following the latter consensus, see the review Adoutte *et al.*, 1999). The intronmapping of NDP kinase seemingly supports the "traditional" phylogenic tree rather than the "rRNA based" tree. Furthermore, recent studies on *C. elegans* genes revealed that the

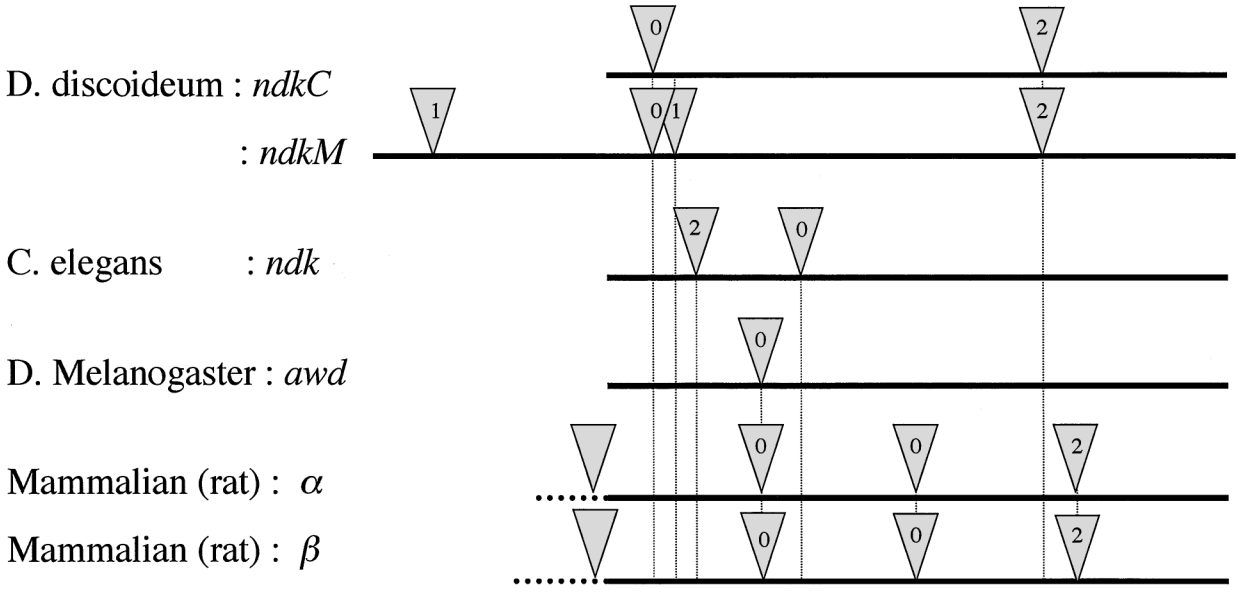

Fig. 3. Intron-mapping of NDP kinase mRNAs from representative species. Triangles indicate positions of the introns. The number in the triangle denotes the phase: 0, intron insertion between two triplet codons; 1, intron insertion after the first base of a codon; 2, intron insertion after the second base. The dotted lines added at each of the 5′-end of the α and the β mRNA indicate the 5′ untranslated sequence.

nematode genes contain the species-specifically located introns (Zang and Maizels, 2001). These data and intron location in the nematode *ndk* strongly suggest that these introns could be inserted in the genome after phylum separation under a species-specific mechanism. However, it cannot be ruled out that the intron of *awd* and those of mammalian ortholog could be inserted at the same position independently, if the "rRNA based" tree were correct. In the case of the two NDP kinase genes of slime mold *ndkC* and *ndkM*, they contain two introns and four introns, respectively. The two intervening sequences of *ndkC* are concordantly located with the second and forth introns of *ndkM* but completely discordant with either of the metazoan species (Fig. 3). The facts also strongly suggest that a single ancestral NDP kinase gene in the slim mold progenitor obtained two (species-specific) intervening sequences after the phylum speciation from a common ancestor of the fungi/plants/animals, followed by its duplication and evolution to the present forms. These intron-mapping data not merely shed the light on the NDP kinase gene evolutionary process *per se*, but also pose intriguing suggestions to universal phylogenic study. Further phylogenic study is needed in order to reach the most probable consensus.

CLUES FOR THE MOLECULAR EVOLUTION OF THE REGULATORY REGION

Since every gene requires a network of other genes in order to exert a biological function, gene expression is a crucial part of gene function. Gene expression is governed by coordinately interacting cis-elements and transactivators. In many bacterial genes, a coregulated system, an operon, regulates a set of functionally associated genes in an orchestrated manner. Interestingly *E. coli ndk* has been reported to generate monocistronic mRNA (Hama *et al.*, 1991). Eucaryotes NDP kinase gene regulation largely remains uncertain. Comparison of the regulatory regions of the orthologous genes reveals another example of the characteristic genomic diversity of the NDP kinase genes, despite that the coding regions are highly conserved. The diversity of the regulatory region might provide the fundamental principles that drive each of these genes to be expressed in a specific spatiotemporal manner, and to exert the pluripotential function of the gene. In particular between phylogenically far distant taxa, the genetic differences of noncoding regions are too large to obtain informative data by simple sequence comparison. In order to avoid such difficulty, some conservative genetic features, such as CpG island and cis-element motifs, are focused on in the analysis of the NDP kinase gene regulatory regions.

CpG dinucleotides cluster (CpG islands) is a characteristically preserved feature found in a constitutively expressed gene regulatory region (Bird, 1986). It provides a useful clue in the search of the active site for the transcriptional initiation. Both NDP kinase genes of the nematode and the fruit fly have a characteristic CpG island that continuously covers the 5′-flanking to the 5′ untranslated

region. The rat α gene and human homolog *nm23-H2* gene also possess a CpG island that is distributed from the 5′ regulatory region to the second exon. In contrast, the β gene and *nm23-H1* demonstrate the pausity of the CpG dinucleotides in the 3′-portion of the first intron. Furthermore there are no sequence homologies between the first intron of the α and β genes. It follows that (1) the α or *nm23-H2* should be representative of the direct descendant of the archetype gene, or ortholog, (2) the β should be a sib-gene, or the nearest paralog, derived from the duplicated α gene ancestor that would be truncated just upstream of the acceptor site of the second exon. Occasionally an acquired alien genomic segment could give the β a new promoter and an untranslated (noncoding) exon, whereas many other paralogous genes would obtain an extra coding exon (or exons) if the alien genomic segments possessed peptide-coding information in the frame of the duplicated NDP kinase gene.

The increase in the number of transcription factor species was reported to be well correlated with the complexity of the body plan (Tupler *et al.*, 2001). It is reasonable to hypothesize that cis-elements for gene regulation also would correlatively evolve with these transcription factors. Concerning NDP kinase genes, the putative regulatory regions have principally increased in terms of size and repertoire of cis-element motifs in correlation with organism complexity. Phylogenic comparison of the regulatory regions based on conserved cis-element motifs revealed intriguing features of the NDP kinase genes (Fig. 4). First, the closely associated E-box motif

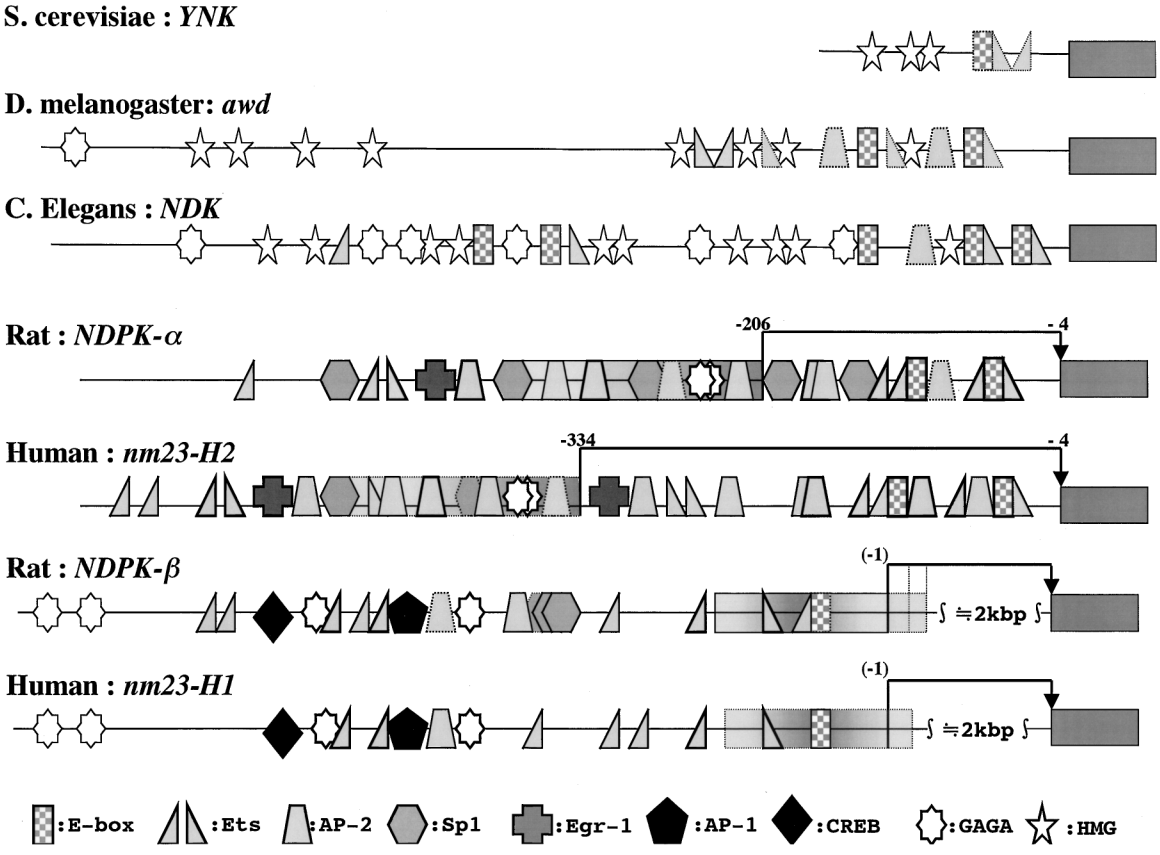

Fig. 4. Comparison of 5′-region of eucaryote NDP kinase genes. Several selected cis-element motifs schematically depicted at the bottom were searched in the 5′-region of the NDP kinase genes derived from representative eucaryote species. Figures demarcated by dotted line indicate the partially degraded motif from the consensus but reported to be potent in literature. Figures drawn by bold line indicate highly conserved motif between homologous genes, that is, the α and *nm23-H2*, and the β and *nm23-H1*, respectively. All the motifs, except Ets, are depicted without considering their orientation, whereas Ets motif (GGAA) on the sense strand is depicted differently from that on template strand, left sidefigure and right side one, respectively. Right side boxes coated in dark gray indicate coding exons, and the boxes painted in graded gray indicate noncoding exons. Numbers in *NDPK-α* and *nm23-H2* indicate relative position from the initiation codon. For description of the position in each *NDPK-β* and *nm23-H1* regulatory region, the major splice donor site of the first exon is used as the starting point (−1).

(CAc/gg/cTG) and Ets consensus motif (c/gaGGAAg) are consistently located adjacent to the translation initiation site in the orthologous genes (*S. cerevisiae ynk*, *D. melanogaster awd*, rat *NDPK-α*, human *nm23-H2*), whereas no such paired consensus exists in the corresponding regions of *NDPK-β* nor *nm23-H1*. Second, the regulatory regions of yeast, the fly, and the nematode are predominantly composed of AT-rich sequences. They contain multiple AT-rich motifs for high mobility group (HMG) binding sites (ATTGTT, CtTTATT,) and the GAGA motif, that would play crucial roles for chromatin remodeling and DNA bending in order to promote transcription, among or in the adjacent upstream of the Ets/E-box motifs. In contrast, the corresponding regions of the *α*, *β*, and the human counterparts bear none of the AT-rich sequences containing the HMG binding sites. Instead of AT-rich sequences, they are composed of rather GC-rich sequences, and contain GC-rich cis-element motifs such as AP-2 binding site motif (gCCcgcGGg), GC-box (gGGGc/tGGGg), and Egr-1 binding site motif (GCGGGGGCG). A majority of these motifs are conservatively located between rat and human homologs (Fig. 4). Third, the rat *α* and human *nm23-H2*, which possess the conserved core sequences including Ets/E-box motifs in their first introns, have more highly conserved genomic segments from around the first exon–intron junction to the upstream region. These segments contain conserved motifs for AP-2, GAGA, Sp1, Egr-1, and Ets, unique to them but distinct from the paralogs (the *β* and *nm23-H1*). Forth, the *β* and *nm23-H1* are distinct from their paralogs the *α* or *nm23-H2* at the 5′ structure. The differences are as follows: (1) the *β* and *nm23-H1* have a relatively longer intervening sequence between the first and second exon, about 2 kbp in length, (2) segmentally preserved sequences between the species in the 5′-regulatory and noncoding first exon, and the highly conserved "stepping stone" like sequences contain several characteristic cis-element motifs (including Ets, AP-2, AP-1, CREB, and GAGA-motif). Fifth, the sequence homology search was performed by using a computer program (Lipman-Pearson method/GENETYX) between homologous regions of the rat *α* and *nm23-H2*, or between those of the rat *β* and *nm23-H1* as follows: (1) within 200 bp sequences upstream from first intron/second exon junctions of the *α* and *nm23-H2*, (2) within 300 bp sequences upstream from first exon/intron junctions of the *α* and *nm23-H2*, and (3) within 300 bp sequences upstream from the first intron/second exon junctions of the *β* and *nm23-H1*. The comparison (including appropriately inserted gaps) provided matching values of (1) 56%, (2) 61%, and (3) 67%, respectively. These values strongly suggest that the degree of the mismatch could occur depending on the "aging process" of these genomic fragments in the NDP kinase genes.

Taken together, we would be allowed to draw a rough draft of the sequentially occurring events in the orthologous NDP kinase genes during animal evolution, as follows: (1) an alien genomic sequence had been inserted in the regulatory region of the orthologous NDP kinase gene in the ancient Deuterostomes ancestor (formation of a "dual" promoter region in the *α* ancestor), (2) at least two gene duplication events occurred in the ortholog: twice in the ortholog and once in its paralog, or three times in the ortholog (formation of Group I gene family), and (3) in the latest gene duplication event a paralogous gene with completely novel regulatory region had arisen in the chordate common ancestor (birth of the *β* or *nm23-H1*). We should consider several factors such as, species-specific base substitution speeds, constraining forces on the regulatory regions, and the long period of a common ancestor during which the genetic element could be interchanged (confer the old allele hypothesis). Further data of other phyla will help us to understand the molecular evolution of the NDP kinase gene, and that would lead to more profound insight into the question how our life evolved.

To confirm the molecular interaction between the putative cis-elements and trans-activator interaction, electrophoretic mobility shift assay (EMSA) was performed (Fig. 5). Most of the selected genomic fragments that contain conservative cis-element motifs between rat and human homologous genes provided binding capacity with certain nuclear protein(s) in the nuclear extracts derived from nonstimulated or nerve growth factor (NGF)-stimulated PC12 cells. In the *α* gene, multiple AP-2, Ets-1, and Sp1 binding sites, and the two E-box sequences showed constitutive binding capacity (no remarkable differences were observed between the data using nonstimulated and NGF-stimulated lysates). In the *β* gene, a couple of Ets-1 and AP-2 binding sites also provided constitutive binding capacity under these conditions. By contrast, AP-1 binding consensus residing in the *β* gene indicated strong responsiveness only to the nuclear extract harvested 1 h after NGF-stimulation (Fig. 5, Lanes 17 and 18). The CREB binding motif also reacted to the stimulated lysate (data not shown). These data demonstrated that the regulatory region of the *β* gene has potently competent elements to the MAP kinase cascade, whereas the *α* gene has constitutive expression at the immediate early phase of the NGF-stimulation.

To address rat NDP kinase *α* gene promoter activity in the cell, CAT assay was performed (Ishikawa *et al.*, 1997). It indicated that one of the strongest core promoter activities seems to reside in the region between −568 and −452 (from the initiation codon position), where the most

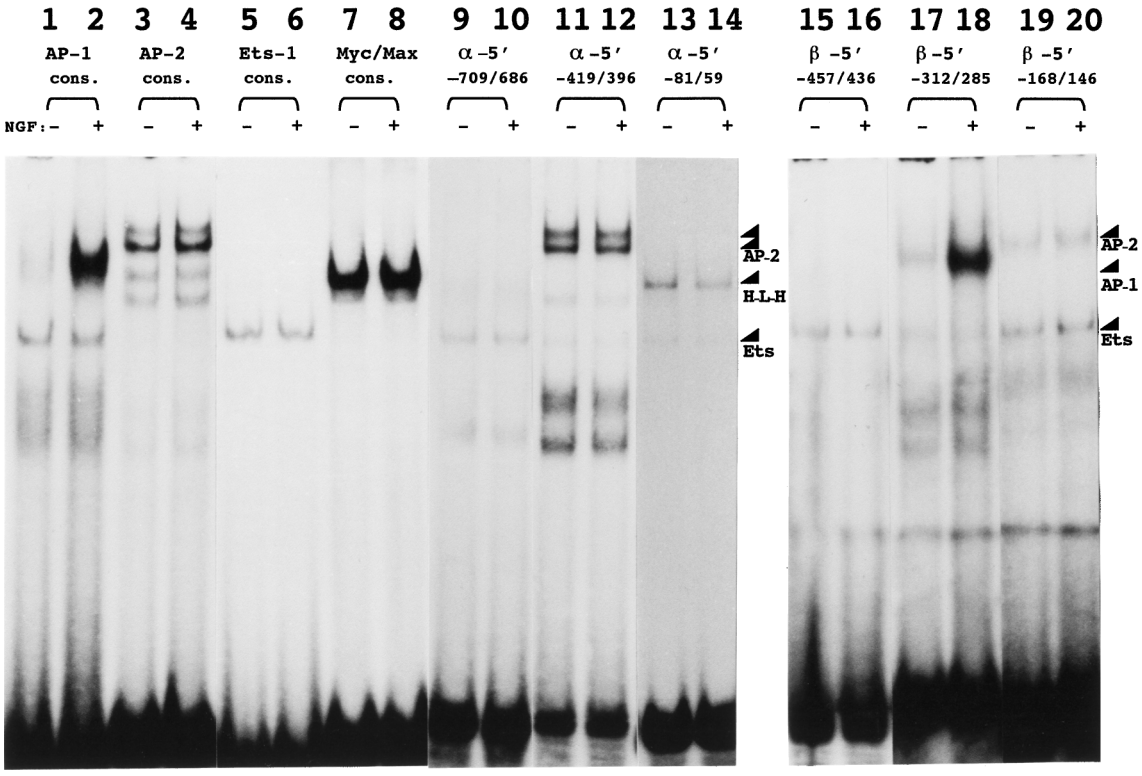

Fig. 5. The molecular interaction between the cis-elements and trans-activators. An autoradiogram displaying the specifically shifted band detected by an electrophoretic mobility shift assay (EMSA). The double stranded oligonucleotides containing the consensus sequence for AP-1, AP-2, Ets-1, and Myc/Max (E-box) were used for controls (Lanes – 1, 2, 3, 4, 5, 6, 7, 8). The oligonucleotides that corresponds to the α and β genes 5′ regulatory region that contain putative cis-element(s) were synthesized. From α gene, three fragments (-709 to -686, -419 to -396, -81 to -59, relative to the translation initiation site), each of which contains motifs for Ets, AP-2 and Egr-1, Ets and E-box, respectively, were prepared. From β gene, three fragments (-457 to -436, -312 to -285, -168 to -146, relative to the major splice donor site), each of which contains motifs for Ets, AP-1 and Ets, AP-2 and Ets, respectively, were used. Nuclear extracts were prepared from PC12D (Ishijima *et al.*, 1999) cells stimulated by NGF for 1 h or nonstimulated ones following previously reported methods (Osborn *et al.*, 1989; Schreiber *et al.*, 1989). EMSA was performed as described previously (Carey and Smale, 2000). Each probe was reacted with the nuclear extract derived from nonstimulated (Lanes – 1, 3, 5, 7, 9, 11, 13, 15, 17, 19) or NGF-stimulated cells (Lanes – 2, 4, 6, 8, 10, 12, 14, 16, 18, 20).

distal putative Sp1 binding consensus sequence GT box, an aberrant form of the GC box, is located (Li *et al.*, 1991). Weak promoter activity was found in the region between -452 and -366. The proximal region from position -366 expresses no significant promoter activity in the fibroblast and a tumor cell line tested. It should be noted that the region showing the strongest promoter activity corresponds to the most distal universally active transcription initiation sites, whereas the region seems to be too far from downstream initiation sites. Considering the "dual" structure of the α gene regulatory region, it could be reasonable to postulate another downstream promoter for these sites. Unfortunately, however, because of the limitation of available host cells that produce the group 2 type transcripts, the downstream promoter activities have not yet been confirmed.

Regarding the promoter analyses of the β or its homologs, about 100 bp stretch upstream from the AP-1 consensus is supposed to be essential for the promoter activity (P. Steeg personal communication). On the other hand, a 350 bp fragment of the 5′ untranslated region of *nm23-M1* was reported to express sufficient promoter activity in the murine ES cells (Debernat *et al.*, 1999). These data are compatible with the highly conservative region among the β homologs.

HETEROGENEOUS mRNAs

Both in the rat α and β genes, numerous transcription initiation sites have been identified in an extraordinarily wide range using the RNAase protection assay and

5′-RACE method (Ishikawa *et al.*, 1992, 1997, and unpublished observation). The α mRNA major transcription initiation sites are located from -30 to -400 (relative to the initiation codon). The mRNA, which starts further upstream than at position -206 is spliced at -206 and accepted at -4, is categorized as group 1 type mRNA. In contrast, mRNA initiated downstream from position -206 is not spliced and bears various sizes of the 5′ untranslated stretch continuing from the cap sites to the translation initiation site at position $+1$. This type of mRNA is categorized as group 2 (Ishikawa *et al.*, 1997). The β mRNA also is initiated over a wide range, about 300 bp in length, in which there are three splice donor sites. Majority of the β mRNA are initiated in upstream of a typical splicing donor site, which is located at one-thirds of this initiation site window, of which sequence shows high conservativeness among mammalian species. However, the rest are initiated over 200 bp stretch downstream of the authentic splice site, where two splice donor sites are exist. Consequently, the β gene generates three different types of mRNA depending on their initiation sites (Fig. 4; Ishikawa *et al.*, unpublished data). In humans, both *nm23-H1* and *H2* also provide multiple initiation sites as in rat genes (Okada *et al.*, 1996). Furthermore, *nm23-H2* generates two types of mRNA species similar to the rat α gene. That is, the initial *nm23-H2* clone (Stahl *et al.*, 1991) is representative of the rat group 1 type α homolog (composed of five exons), whereas the *PuF* clone (Postel *et al.*, 1993) belongs to the group 2 type homolog (composed of four exons).

Heterogeneous transcription initiation can be interpreted mainly from two points of view: Different transcription initiation sites (1) make different peptides products, and (2) are regulated under distinct elements. Regarding the rat NDP kinase isoforms, despite the huge heterogeneity of the 5′-regions, there exist stop codons between several upstream ATGs and the authentic initiation codon of the enzyme, thus ruling out the former possibility. On the other hand, the RNAase protection analyses and *in situ* hybridization method demonstrated differential expression of the groups 1 and 2 types of the α mRNA in tissues and cell lines (Fig. 1(c–e)). The upstream regulatory region might play an essential role in constitutive and predominant expression of the group 1 type α mRNA in most cells, whereas the putative downstream regulatory region could be responsible for the expression of group 2 type α mRNA in certain tissues or cells such as muscles and gastric gland cells. *In situ* hybridization analyses provided informative data in support of this idea: (1) predominant expression of group 1 mRNA was confirmed in most cells examined, (2) group 2 mRNA was expressed in some of these cells, there were no cell types that exhibited group 2 mRNA alone, and (3) it should be noted that in some cells only group 1 mRNA was almost exclusively expressed. It follows that, in cooperation with the universal core promoter activity, additional promoter activities could be generated by downstream cis-elements and trans-acting factors, such as putative E-box consensus sequences and the muscle-specific transcription factors (reviewed in Schwarz *et al.*, 1993) for the preferential induction of the group 2 transcripts in muscle cells.

POSTTRANSCRIPTIONAL REGULATION

Although numerous different forms of transcripts for the α isoform have been defined, their physiological meaning is largely unknown. We have examined several possibilities, including structural polymorphism at the peptide level for both types of the α gene products. But so far these attempts have been unsuccessful, except in vitro translation analyses, in which synthetic RNAs generated from each representative of the groups 1 and 2 types of α cDNA clones and the β cDNA clone revealed remarkable differences in translation efficiency; the group 2 type α RNA was most effectively translated among them, while the group 1 type α RNA was minimal. Relative efficiency rates were approximately 100, 34, and 14%, respectively. Regarding translational control mechanisms (reviewed in Hershey, 1991), no definitive data have been revealed that indicate regulatory mechanisms of NDP kinase expression at the posttranscriptional steps. The data suggest the possible biological significance of the multiple forms of the 5′ untranslated region; group 1 type α mRNA, which is constitutively expressed, may work for an "idling" state at a minimal translating rate, whereas the "induced or lineage-specific" form (group 2 type mRNA) could be used at a higher rate in case of need. Hence these differential translation rates could guarantee homeostasis of triphosphate nucleotide pools in different circumstances. The predicted conservation of these multiple forms of NDP kinase mRNA for the α isoform among mammalian species could also give a rationale for its biological significance.

The birth of the multicellular organism was a revolutionary event in the long history of life. In order to elucidate what forces or mechanisms promote the generation of the multicellular organism is a still unsolved question. Interesting, drastically decreased production of the NDP kinase protein was observed in slime mold during development of a multicellular organization (Wallet *et al.*, 1990). Similar phenomenon was observed in *M. xanthus* during a multicellular aggregate formation (Muñoz-Dorado *et al.*, 1990). Furthermore, the regulation of the NDP kinase protein contents differs sharply between unicellular organisms and multicellular organisms. In unicellular

organisms, overexpression of the enzyme was achieved easily to a vast extent. For example more than 100 times overexpression in *E. coli* and *S. cerevisea* were observed, whereas it has hardly succeeded in higher eucaryotes in spite of laborious attempts. At the very early stage of the cloning and characterization of the rat NDP kinase mRNA, it was noticed that the quantities of mRNA are not directly related to those of the protein or to enzymatic activity; the NDP kinase mRNA levels vary more than 10-fold among different organs tested, whereas levels of the protein or enzymatic activity vary at most by 2-fold (Kimura *et al.*, 1990). Similar enigmatic phenomena were also observed in regard to cell immortalization: the mRNA was increased severalfold, whereas the protein amount increased only at most 1.5-fold (Ohneda *et al.*, 1994). These observations suggest possible regulations of the NDP kinase protein level by posttranslation regulation mechanism.

A PERSPECTIVE: GENE REGULATION AND MULTIPLE FUNCTION OF NDP KINASE

The expression pattern of the rat NDP kinase α and β mRNA demonstrated that these major genes are expressed ubiquitously but in a tissue/cell lineage-specific manner (Fig. 1(c–e)). The protein level of the rat NDP kinase also indicates that major isoforms accumulate ubiquitously but in a stage-specific and/or tissue-specific manner (Fig. 1(g) and (h)). Most of the conserved sequences in the regulatory regions in the α and the β genes contain putative cis-element motifs, and provide binding capacities to the estimated nuclear proteins (Fig. 5). Taken together, it can be considered that the expression of NDP kinase at the transcription level is governed by "multiple-layered" constructs of regulatory regions, each of which contains various motifs for the constitutive expression, for stage/lineage-specific expression, and for responsive elements to various stimuli. The highly integrated system is embodied in the mammalian orthologs (the rat α, *nm23-H2*) and their paralogs (the rat β, *nm23-H1*). The huge heterogeneity of their mRNA at 5′ untranslated region due to multiple initiation sites would generate another layer of gene regulation: regulation at posttranscriptional (translational) steps. Further studies, including phylogenic comparison and precise examination of the molecular interaction of cis-elements and the trans-activators, are necessary in order to gain a complete understanding of the interwoven complexity of gene regulation. In regard to gene regulation function, the recently proposed nuclease function of the major NDP kinase isoform (nm23-H2/PuF; Postel, 1999) has provided the possible molecular basis for the transcriptional regulation of some genes. Postel reviewed the novel property of NDP kinase in the present issue. How and when this function would have been acquired during NDP kinase molecular evolution is an intriguing question.

The fact that the gross configuration at the peptide level is highly conserved from eubacteria to mammalia demonstrates a gradual changing process of NDP kinase proteins during evolution. However, we can recognize a couple of "phase transition"-like drastic changes during the molecular evolution of this pivotal enzyme. These would include oligomerization in bacterial form (quadrimer) to eucaryote form (hexamer), loss of NDP kinase during parasitic adaptation of mycoplasma, and the constrained protein dose in multicellular organisms, gain and loss of the large fragments in the regulatory region. Further studies from the viewpoint of molecular evolution are also essential to an understanding of these phenomena and NDP kinase function.

Note

The draft genome sequence of sea squirt *Ciona intestinalis* has been reported (Dehal *et al.*, 2002, Science **298**, 2157–2167). The putative orthologs of *nm23-H1*, *H2*, *H5*, *H6*, and *H8* have been annotated in the genome.

REFERENCES

Adams, M. D. *et al.* (2000). *Science* **287**, 2185–2195. (Gene Bank accession no. AE003779 (gene= awd))

Adoutte, A., Balavoine, G., Lartillot, N., and de Rosa, R. (1999). *Trends Genet.* **15**, 104–108.

Arnaud-Dabernat, S., Bourbon, P. M., Dierich, A., La Meur, M., and Daniel, J.-Y. (2003). *J. Bioenerg. Biomembr*. **35**, 19–30.

Biggs, J., Hersperger, H., Steeg, P. S., Liotta, L. A., and Shearn, A. (1990). *Cell* **63**, 933–940.

Biggs, J., Tripoulas, N., Hersperger, E., Dearolf, C., and Shearn, A. (1988). *Genes Dev.* **2**, 1333–1343.

Bird, A. P. (1986). *Nature* (*London*) **321**, 209–213.

Bult, C. J., White, O., Olsen, G. J., *et al.* (1996). *Science* **273,** 1058–1073.

Carey, M., and Smale, S. T. (2000). *Transcriptional Regulation in Eukaryotes*, Cold Spring Harbor Laboratory Press, New York.

Cavalier-Smith, T. (1985). *Nature* **315**, 283–284.

Dabernat, S., Larou, M., Massé, K., Dobremez, E., Landry, M., Mathieu, C., and Daniel, J.-Y. (1999). *Gene* **236**, 221–230.

Dearolf, C. R., Tripoulas, N., Biggs, J., and Shearn, A. (1988). *Dev. Biol.* **129**, 169–178.

De La Rosa, A., Williams, R. L., and Steeg, P. S. (1995). *Bioessays* **17,** 53–62.

Doolittle, R. F., Feng, D.-F., Tsang, S., Cho, G., and Little, E. (1996). *Science* **271**, 470–477.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison III, C. A., and Venter, J. C. (1995). *Science* **270**, 397–403.

Fukuchi, T., Nikawa, J., Kimura, N., and Watanabe, K. (1993). *Gene* (*Amst.*) **129**, 141–146.

Fukuda, M., Ishii, A., Yasutomo, Y., Shimada, N., Ishikawa, N., Nagata, N., Irimura, T., Nicolson, G. L., and Kimura, N. (1996). *Int. J. Cancer* **65**, 531–537.

Gilbert, W. (1987). *Cold Spring Harb. Symp. Quant. Biol.* **52**, 901–905.

Gilles, A.-M., Presecan, E., Vonica, A., and Lascu, I. (1991). *J. Biol. Chem.* **266**, 8784–8789.

Glass, J. I., Lefkowitz, E. J., Glass, J. S., Heiner, C. R., Chen, E. Y., and Cassell, G. H. (2000). *Nature* (*London*), **407**, 757–762.

Hama, H., Almaula, N., Lerner, C. G., Inouye, S., and Inouye, M. (1991). *Gene* (*Amst.*) **105**, 31–36.

Hershey, J. W. B. (1991). *Annu. Rev. Biochem.* **60**, 717–755.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.-C., and Herrmann, R. (1996). *Nucleic Acids Res.* **24**, 4420–4449.

Ishijima, Y., Shimada, N., Fukuda, M., Miyazaki, H., Orlov, N. Y., Orlova, T. G., Yamada, T., and Kimura, N. (1999). *FEBS Lett.* **445**, 155–159.

Ishikawa, N., Shimada, N., Munakata, Y., Watanabe, K., and Kimura, N. (1992). *J. Biol. Chem.* **267**, 14366–14372.

Ishikawa, N., Taniguchi-Seto, H., Munakata, Y., Takagi, Y., Shimada, N., and Kimura, N. (1997). *J. Biol. Chem.* **272**, 3286–3295.

Izumiya, H., and Yamamoto, M. (1995). *J. Biol. Chem.* **270**, 27859–27864.

Keim, D., Hailat, N., Melhem, R., Zhu, X. X., Lascu, I., Veron, M., and Strahler, J. (1992). *J. Clin. Invest.* **89**, 919–924.

Kimura, N., Shimada, N., Nomura, K., and Watanabe, K. (1990). *J. Biol. Chem.* **265**, 15744–15749.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessierès, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S.-K., Codani, J.-J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E, Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V. , Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B.-S., Soldo, B., Sorokin A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M ., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K ., Yoshikawa, H.-F., Zumstein, E., Yoshikawa, H., and Danchin, A. (1997). *Nature* **390**, 249–256.

Lacombe, M.-L., Millon, L., Munier, A., Mehus, J. G., and Lambeth, D. O. (2000). *J. Bioener. Biomembrg.* **32**, 247–258.

Lacombe, M.-L., Wallet, V., Troll, H., and Veron, M. (1990). *J. Biol. Chem.* **265**, 10012–10018.

Lakso, M., Steeg, P. S., and Westphal, H. (1992). *Cell Growth Differ.* **3**, 873–879.

Leone, A., Flatow, U., King, C. R., Sandeen, M. A., Margulies, I. M. K., Liotta, L. A., and Steeg, P. S. (1991). *Cell* **65**, 25–35.

Li, R., Knight, J. D., Jackson, S. P., Tijan, R., and Botchan, M. R. (1991). *Cell* **65**, 493–505.

Lu, G., Zhang, X., Almaula, N., Mathews, C. K., and Inouye, M. (1995). *J. Mol. Biol.* **254**, 337–341.

Muñoz-Dorado, J., Inouye, M., and Inouye, S. (1990). *J. Biol. Chem.* **265**, 2702–2706.

Ohneda, K., Fukuda, M., Shimada, N., Ishikawa, N., Ichou, I., Kaji, K., Toyota, T., and Kimura, N. (1994). *FEBS Lett.* **348**, 273–277.

Ohno, S. (1967). *Sex Chromosomes and Sex-Linked Genes*, Springer, Berlin.

Okada, K., Urano, T., Baba, H., Furukawa, K., Furukawa, K., and Shiku, H. (1996). *Oncogene* **13**, 1937–1943.

Osborn, L., Kunkel, S., and Nabel, G. (1989). *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2336–2340.

Postel, E. H. (1999). *J. Biol. Chem.* **274**, 22821–22829.

Postel, E. H., Berberich, S. J., Flint, S. J., and Ferrone, C. A. (1993). *Science* **261**, 478–480.

Schopf, J. W. (1998). In *The Molecular Origins of Life: Tracing the Roots of the Universal Tree of Life* (Brack, A., ed.), Cambridge University Press, Cambridge, England, pp. 336–362.

Schreiber, E., Matthias, P., Muller, M., and Schaffner, W. (1989). *Nucleic Acids Res.* **17**, 6419.

Schwarz, J. J., Martin, J. F., and Olson, E. N. (1993). In *Gene Expression: General and Cell Type Specific* (Karin, M., ed.), Birkhauser, Boston, pp. 93–115.

Shimada, N., Ishikawa, N., Munakata, Y., Toda, T., Watanabe, K., and Kimura, N. (1993). *J. Biol. Chem.* **268**, 2583–2589.

Stahl, J. A., Leone, A., Rosengard, A. M., Porter, L., King, C. R., and Steeg, P. S. (1991). *Cancer Res.* **51**, 445–449.

Steeg, P. S., Bevilacqua, G., Kopper, L., Thorgeirsson, U. P., Talmadge, J. E., Liotta, L. A., and Sobel, M. E. (1988). *J. Natl. Cancer Inst.* **80**, 200–204.

The *C. elegans* Sequencing Consortium. (1998). *Science* **282**, 2012. (Gene Bank accession no. Z79754 (gene=F25H2.5))

Troll, H., Winckler, T., Lascu, I., Muller, N., Saurin, W., Veron, M., and Mutzel, R. (1993). *J. Biol. Chem.* **268**, 25469–25475.

Trotman, C. N. A. (1998). *Trends Genet.* **14**, 132–134.

Tsuiki, H., Nitta, M., Furuya, A., Hanai, N., Fujiwara, T., Inagaki, M., Kochi, M., Ushio, Y., Saya, H., Nakamara, H. (1999). *J. Cell. Biochem.* **76**, 254–269.

Tupler, R., Perini, G., and Green, M. R. (2001). *Nature* **409**, 832–833.

Urano, T., Takamiya, K., Furukawa, K., and Shiku, H. (1992). *FEBS Lett.* **309**, 358–362.

Wallet, V., Mutzel, R., Troll, H., Barzu, O., Wurster, B., Veron, M., and Lacombe, M.-L. (1990). *J. Natl. Cancer Inst.* **82**, 1199–1202.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., and Koonin, E. V. (2002). *Trends Genet.* **18**, 472–478.

Zang, X., and Maizels, R. M. (2001). *Trends Biochem. Sci.* **26**, 191–197.