

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,200

Open access books available

116,000

International authors and editors

125M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Understanding Protein Function - The Disparity Between Bioinformatics and Molecular Methods

Katarzyna Hupert-Kocurek<sup>1</sup> and Jon M. Kaguni<sup>2</sup>

<sup>1</sup>University of Silesia

<sup>2</sup>Michigan State University

<sup>1</sup>Poland

<sup>2</sup>United States of America

## 1. Introduction

Bioinformatics has its origins in the development of DNA sequencing methods by Alan Maxam and Walter Gilbert (Maxam and Gilbert, 1977), and by Frederick Sanger and coworkers (Sanger et al., 1977). By entirely different approaches, the first genomes determined at the nucleotide sequence level were that of bacteriophage  $\phi$ X174, and the recombinant plasmid named pBR322 composed of about 5,400 (Sanger et al., 1977), or 4,400 base pairs (Sutcliffe, 1979), respectively. In contrast, two articles that appeared in February 2001 reported on the preliminary DNA sequence of the human genome, which corresponds to 3 billion nucleotides of DNA sequence information (Lander et al., 2001; Venter et al., 2001). Only two years later, the GenBank sequence database contained more than 29.3 billion nucleotide bases in greater than 23 million sequences. With the development of new technologies, experts predict that the cost to sequence an individual's DNA will be about \$1000. This reduction in cost suggests that efforts in the area of comparative genomics will increase substantially, leading to an enormous database that vastly exceeds the existing one. By way of comparative genomics approaches, computational methods have led to the identification of homologous genes shared among species, and their classification into superfamilies based on amino acid sequence similarity. In combination with their evolutionary relatedness, superfamily members have been clustered into clades. In addition, high throughput sequencing of small RNAs and bioinformatics analyses have contributed to the identification of regions between genes that can code small RNAs (siRNA, microRNA, and long noncoding RNA), which act during the development of an organism to modulate gene expression at the post-transcriptional level (Fire et al., 1998; Hamilton and Baulcombe, 1999) reviewed in Elbashir et al., 2001; Ghildiyal and Zamore, 2009; Christensen et al., 2010). An emerging area is functional genomics whereby gene function is deduced using large-scale methods by identifying the involvement of specific genes in metabolic pathways. More recently, phenotype microarray methods have been used to correlate the functions of genes of microbes with cell phenotypes under a variety of growth conditions (Bochner, 2009). These methods contrast with the traditional approach of mapping a gene via the phenotype of a mutation, and deducing the function of the gene product based on its biochemical analysis in concert with physiological studies. Such studies have been performed to confirm the functional importance of conserved residues shared by superfamily members, and also

to determine the role of specific residues for a given protein. In comparison, comparative genomics methods are unable to distinguish if a nonconserved amino acid among superfamily members is functionally important, or simply reflects sequence divergence due to the absence of selection during evolution. Without functional information, it is not possible to determine if a nonconserved amino acid is important.

## 2. Bioinformatics analysis of AAA+ proteins

On the basis of bioinformatics analysis, P-loop nucleotide hydrolases compose a very large group of proteins that use an amino acid motif named the phosphate binding loop (P-loop) to hydrolyze the phosphate ester bonds of nucleotides. A positively charged group in the side chain of an amino acid (often lysine) in the P-loop promotes nucleotide hydrolysis by interacting with the phosphate of the bound nucleotide. Additional bioinformatics analysis of this group of proteins led to a category of nucleotidases containing the Walker A and B motifs, as well as additional motifs shared by the AAA (ATPases Associated with diverse cellular Activities) superfamily (Beyer, 1997; Swaffield and Purugganan, 1997). These diverse activities include protein unfolding and degradation, vesicle transport and membrane fusion, transcription and DNA replication. The additional motifs of the AAA superfamily differentiate its members from the larger set of P-loop nucleotidases. Neuwald *et al.*, and Iyer *et al.* then integrated structural information with bioinformatics analysis to classify members of the AAA+ superfamily into clades (Neuwald *et al.*, 1999; Iyer *et al.*, 2004). These clades are the clamp loader clade, the DnaA/CDC6/ORC clade, the classical AAA clade, the HslU/ClpX/Lon/ClpAB-C clade, and the Helix-2 insert clade. The last two clades have been organized into the Pre-sensor 1 hairpin superclade.

Members of the superfamily of AAA+ ATPases carry a nucleotide-binding pocket called the AAA+ domain that ranges from 200 to 250 amino acids, which is formed by an  $\alpha\beta\alpha$ -type Rossmann fold followed by several  $\alpha$  helices (Figure 1) (Lupas and Martin, 2002; Iyer *et al.*, 2004; Hanson and Whiteheart, 2005). Such proteins often assemble into ring-shaped or helical oligomers (Davey *et al.*, 2002; Iyer *et al.*, 2004; Erzberger and Berger, 2006). Using the nomenclature of Iyer *et al.*, the Rossmann fold is formed by a  $\beta$  sheet of parallel strands arranged in a  $\beta 5$ - $\beta 1$ - $\beta 4$ - $\beta 3$ - $\beta 2$  series. Its structure resembles a wedge. An  $\alpha$  helix preceding the  $\beta 1$  strand and a loop that is situated across the face of the  $\beta$  sheet is a distinguishing feature of the AAA+ superfamily. Another characteristic is the position of several  $\alpha$  helices positioned above the wide end of the wedge. The P-loop or the Walker A motif (GX<sub>4</sub>GKT/S where X is any amino acid) is located between the  $\beta 1$  strand and the following  $\alpha$  helix. The Walker B motif ( $\phi\phi\phi\phi$ DE where  $\phi$  is a hydrophobic amino acid) coordinates a magnesium ion complexed with the nucleoside triphosphate via the conserved aspartate residue. The conserved glutamate is thought to interact with a water molecule to make it a better nucleophile for nucleotide hydrolysis.

AAA+ proteins also share conserved motifs named the Sensor 1, Box VII, and Sensor 2 motifs that coordinate ATP hydrolysis with a change in conformation (Figure 1) (Lupas and Martin, 2002; Iyer *et al.*, 2004; Hanson and Whiteheart, 2005). Relative to the primary amino acid sequence, these motifs are on the C-terminal side of the Walker B motif. The Sensor 1 motif contains a polar amino acid at the end of the  $\beta 4$  strand. On the basis of the X-ray crystal structure of N-ethylmaleimide-sensitive factor, an ATPase involved in intracellular vesicle fusion (Beyer, 1997; Swaffield and Purugganan, 1997), this amino acid together with

the acidic residue in the Walker B motif interacts with and aligns the activated water molecule during nucleotide hydrolysis. The Box VII motif, which is also called the SRH (Second Region of Homology) motif, contains an arginine named the arginine finger by its analogous function with the corresponding arginine of GTPase activator proteins that interacts with GTP bound to a small G protein partner to promote GTP hydrolysis. The crystal structures of several AAA+ proteins have shown that the Box VII motif in an individual molecule is located some distance away from the nucleotide binding pocket. In AAA+ proteins that assemble into ring-shaped or helical oligomers, the Box VII motif of one protomer directs an arginine residue responsible for interaction with the  $\gamma$  phosphate of ATP toward the ATP binding pocket of the neighboring molecule. It is proposed that this interaction or lack thereof coordinates ATP hydrolysis with a conformational change. The Sensor 2 motif, which resides in one of the  $\alpha$  helices that follow the Rossmann fold, also contains a conserved arginine. For proteins whose structures contain the bound nucleoside triphosphate or a nucleotide analogue, this amino acid interacts with the  $\gamma$  phosphate of the nucleotide. As reviewed by Ogura (Ogura et al., 2004), this residue is involved in ATP binding or its hydrolysis in some but not all AAA+ proteins. Like the arginine finger residue, this arginine is thought to coordinate a change in protein conformation with nucleotide hydrolysis.

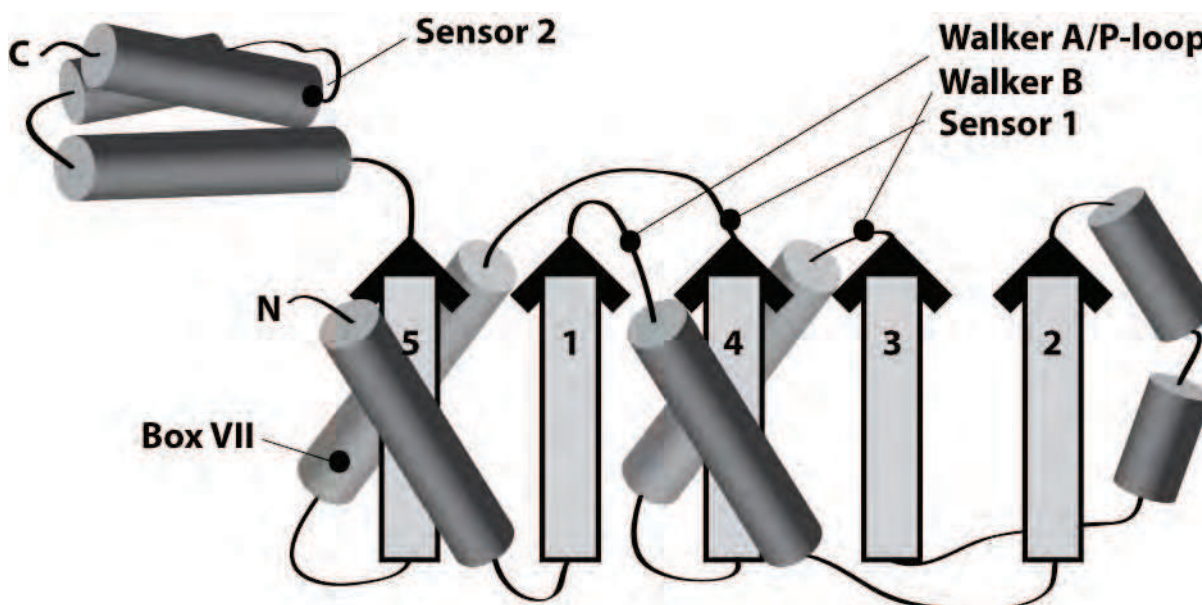


Fig. 1. Structural organization of the AAA+ domain, and the locations of the Walker A/P-loop, Walker B, Sensor 1, Box VII and Sensor 2 motifs are shown (adapted from ref. (Erzberger and Berger, 2006)).

Because this chapter focuses on members of the DnaA/CDC6/ORC or initiator clade, the following summarizes properties of this clade and not others. Like the clamp loader clade, proteins in the initiator clade as represented by DnaA and DnaC have a structure resembling an open spiral on the basis of X-ray crystallography (Erzberger et al., 2006; Mott et al., 2008). In comparison, oligomeric proteins in the remaining clades form closed rings. A characteristic feature of proteins in the initiator clade is the presence of two  $\alpha$  helices between the  $\beta 2$  and  $\beta 3$  strands (Figure 1). Compared with the function of DnaA in the initiation of *E. coli* DNA replication, DnaC plays a separate role. Their functions are

described in more detail below. The ORC/CDC6 group of eukaryotic proteins in the initiator clade, like DnaA and DnaC, act to recruit the replicative helicase to replication origins at the stage of initiation of DNA replication (Lee and Bell, 2000; Liu et al., 2000). The origin recognition complex (ORC) is composed of six related proteins named Orc1p through Orc6p, and likely originated along with Cdc6p from a common ancestral gene.

Bioinformatics analysis of DnaC suggests that this protein is a paralog of DnaA, arising by gene duplication and then diverging with time to perform a separate role from DnaA during the initiation of DNA replication (Koonin, 1992). This notion leads to the question of what specific amino acids are responsible for the different functions of DnaA and DnaC despite the shared presence of the AAA+ amino acid sequence motifs. Presumably, specific amino acids that are not conserved between these two proteins have critical roles in determining their different functions, but how are these residues identified and distinguished from those that are not functionally important? In addition, some amino acids that are conserved among homologous DnaC proteins, which were identified by multiple sequence alignment of twenty-eight homologues (Figure 2), are presumably responsible for the unique activities of DnaC, but what are these unique activities? These issues underscore the limitation of deducing the biological function of protein by relying only on bioinformatics analysis.

### 3. Reverse genetics as an approach to identify the function of an unknown gene

Using various amino acid sequence alignment methods for a particular gene, the postulated function for this gene remains unknown if amino acid sequence homology is not obtained relative to a gene of known function. In such cases, the general approach is to employ reverse genetics to attempt to correlate a phenotype with a mutation in the gene. By way of comparison, forward genetics begins with a phenotype caused by a specific mutation at an unknown site in the genome. The approximate position of the gene can be determined by classical genetic methods that involve its linkage to another mutation that gives rise to a separate phenotype. Refined linkage mapping can localize the gene of interest, followed by PCR (polymerase chain reaction) amplification of the region and DNA sequence analysis to determine the nature of the mutation. As a recent development, whole genome sequencing has been performed to map mutations, dispensing with the classical method of genetic linkage mapping (Lupski et al., 2010; Ng and Kirkness, 2010). The DNA sequence obtained may reveal that the gene and the corresponding gene product have been characterized in the same or different organism, and disclose its physiological function.

In a reverse genetics approach with a haploid organism, the standard strategy is to inactivate the gene with the hope that a phenotype can be measured. Inactivation can be achieved either by deleting the gene or by insertional mutagenesis, usually with a transposon. As examples, transposon mutagenesis has been performed with numerous microbial species, and with *Caenorhabditis elegans* (Vidan and Snyder, 2001; Moerman and Barstead, 2008; Reznikoff and Winterberg, 2008). Using *E. coli* or *S. cerevisiae* as model organisms for gene disruption, one method relies on replacing most of the gene with a drug resistance cassette, or a gene that causes a detectable phenotype. The technique of gene disruption relies on homologous recombination in which the drug resistance gene, for example, has been joined to DNA sequences that are homologous to the ends of the target gene (Figure 3). After introduction of this DNA into the cell, recombination between the ends of the transfected DNA and the homologous regions in the chromosome leads to

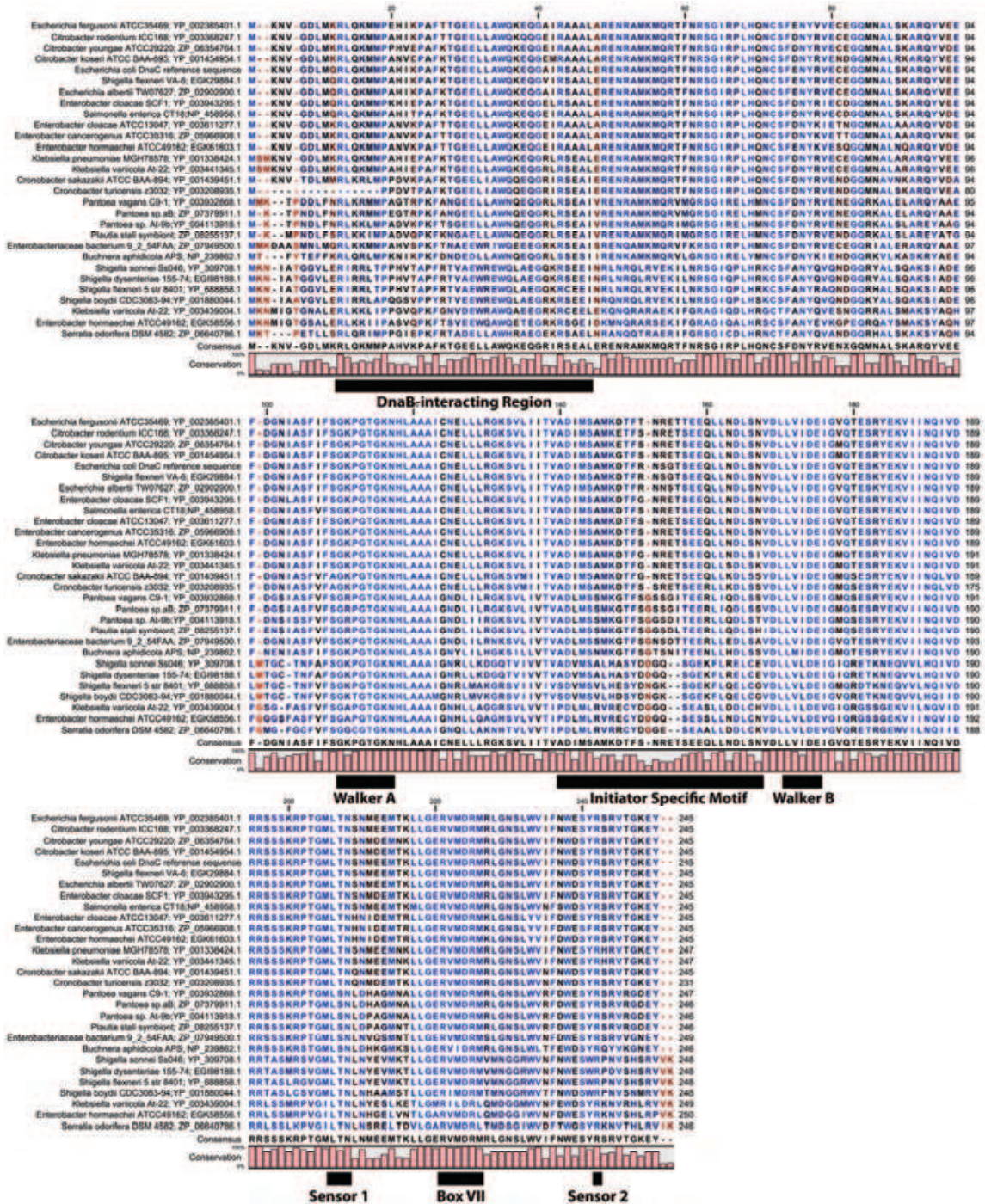


Fig. 2. Multiple sequence alignment of *dnaC* homologues using the Constraint-based Multiple Alignment Tool reveals amino acids that are highly conserved. The Walker A, Walker B, Sensor I, Sensor II and Box VII motifs shared among the AAA+ family of ATPases are shown below the alignment. These motifs are involved in ATP binding, ATP hydrolysis, and coordinating a conformational change with ATP hydrolysis. The figure also shows the region of DnaC near the N-terminus that is involved in interaction with DnaB helicase (Ludlam et al., 2001; Galletto et al., 2003; Mott et al., 2008), and the ISM (Initiator/loader-Specific Motif), which corresponds to two  $\alpha$  helices between the  $\beta$ 2 and  $\beta$ 3 strands, that distinguishes members of the DnaA/CDC6/ORC clade from other AAA+ clades. The ISM causes DnaC assembled as oligomers to form a spiral structure (Mott et al., 2008).

replacement of the chromosomal copy of the gene with the drug resistance cassette, after which the excised copy of the chromosomal gene is lost. In both *E. coli* and *S. cerevisiae*, this approach has been used in seeking to correlate a phenotype with genes of unknown function, and to identify those that are essential for viability (Winzeler et al., 1999; Baba et al., 2006). By either gene disruption or transposon mutagenesis, genetic mapping of the mutation can be performed by inverse PCR where primers complementary to a sequence near the ends of the drug resistance cassette or the transposon are used. This approach first involves partially digesting the chromosomal DNA with a restriction enzyme followed by ligation of the resulting fragments to form a collection of circular DNAs. DNA sequence analysis of the amplified DNA with the primers described above followed by comparison of the nucleotide sequence with the genomic DNA sequence can identify the site of the disrupted gene, or the site of insertion of the transposon.

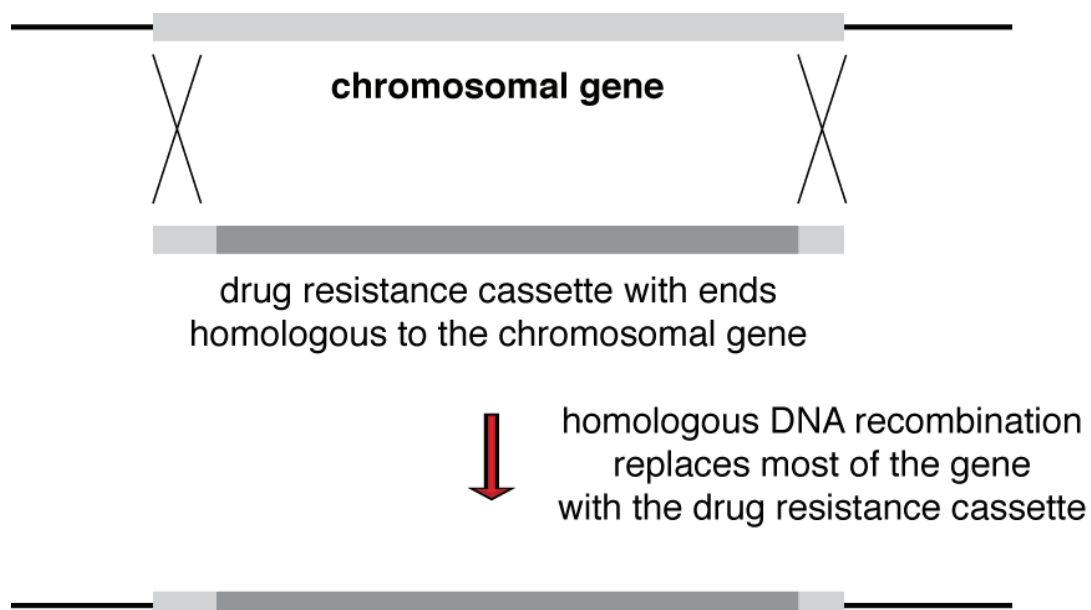


Fig. 3. DNA recombination between the chromosomal gene and homologous DNA at ends of the drug resistance gene leads to replacement of the chromosomal copy of the gene with the drug resistance cassette. The chromosomal gene, and homologous DNA at the ends of the drug resistance cassette are indicated by the lighter shaded rectangles. The drug resistance gene is indicated by the darker rectangles. The thin line represents chromosomal DNA flanking the gene.

With a multicellular organism, a similar strategy that relies on homologous recombination is used to delete a gene. The type of cell to introduce the deletion is usually an embryonic stem cell so that the effect of deletion can be measured in the whole organism. Many eukaryotic organisms have two complete sets of chromosomes. Because the process of homologous recombination introduces the deletion mutation in one of the two pairs of chromosomes, yielding a heterozygote, the presence of the wild type copy on the sister chromosome may conceal the biological effect of the deletion. Thus, the ideal objective is to delete both copies of a gene in order to measure the associated phenotype. To attempt to obtain an organism in which both copies of a gene have been “knocked out,” the standard strategy is to mate heterozygous individuals. By Mendelian genetics, one-fourth of the progeny should carry the deletion in both copies of the gene. The drawback with the approach of deleting a gene

is that it may be essential for viability as suggested if a homozygous knockout organism cannot be obtained. Another pitfall is that it may not be possible to construct a heterozygous knockout because the single wild type copy is insufficient to maintain viability. In either case, no other hint of gene function is obtained except for the requirement for life.

Another complication with attempting to determine the role of a eukaryotic gene by deleting it is the existence of gene families where a specific biochemical function is provided by allelic variants. Hence, to correlate a phenotype by introducing a mutation into a specific allelic variant requires inactivation of all other members of the family. A further complication with eukaryotic organisms is that a product formed by an enzyme of interest in one biochemical pathway may be synthesized via an alternate pathway that involves a different set of proteins. In these circumstances, deletion of the gene does not yield a measurable phenotype.

In the event that deletion of a gene is not possible, an alternate approach to characterize the function of an unknown gene is by RNA interference (reviewed in Carthew and Sontheimer, 2009; Fischer, 2010; Krol et al., 2010). This strategy exploits a natural process that acts to repress the expression of genes during development, or as cells progress through the cell cycle (Fire et al., 1998; Ketting et al., 1999; Tabara et al., 1999). Small RNA molecules named microRNA (miRNA) and small interfering RNA (siRNA) become incorporated into a large complex called the RNA-inducing silencing complex (RISC), which reduces the expression of target genes by facilitating the annealing of the RNA with the complementary sequence in a messenger RNA (Liu et al., 2003). The duplex RNA is recognized by a component of the RISC complex, followed by degradation of the messenger RNA to block its expression. The RNA interference pathway has been adapted as a method to reduce or “knockdown” the expression of a specific gene in order to explore its physiological function. Compared with other genetic methods that examine the effect of a specific amino acid substitution on a particular activity of a multifunctional protein, the knockout and knockdown approaches are not as refined in that they measure the physiological effect of either the reduced function, or the loss of function of the entire protein.

#### **4. *E. coli* as a model organism for structure-function studies**

*Escherichia coli* is a rod-shaped bacterium (0.5 micron x 2 microns in the nongrowing state) that harbors a  $4.4 \times 10^6$  base pair genome encoding more than 4,000 genes. By transposon-based insertional mutagenesis and independently by systematic deletion of each open reading frame, these genes have been separated into those that are essential for viability, and those that are considered nonessential (Baba et al., 2006). Of the total, about 300 genes are of undetermined function, including 37 genes that are essential. BLAST analysis indicates that some of the genes of unknown function are conserved among bacteria, suggesting their functional importance.

In comparison, many of the genes of known function have been studied extensively. Among these are the genes required for duplication of the bacterial chromosome, including a subset that acts at the initiation stage of DNA replication. The following section describes a specific example that focuses on DnaC protein. Studies on this protein take advantage of bioinformatics in combination with its X-ray crystallographic structure, molecular genetic analysis, and the biochemical characterization of specific mutant DnaC proteins to obtain new insight into its role in DNA replication.



## 5. Molecular analysis of *E. coli* DnaC, an essential protein involved in the initiation of DNA replication, and replication fork restart

DNA replication is the basis for life. Occurring only once per cell cycle, DNA replication must be tightly coordinated with other major cellular processes required for cell growth so that each progeny cell receives an accurate copy of the genome at cell division (reviewed in DePamphilis and Bell, 2010). Improper coordination of DNA replication with cell growth leads to aberrant cell division that causes cell death in severe cases. In addition, the failure to control the initiation process leads to excessive initiations, followed by the production of double strand breaks that apparently arise due to head-to-tail fork collisions. In eukaryotes, aneuploidy and chromosome fusions appear if the broken DNA is not fixed that can lead to malignant growth.

In bacteria, chromosomal DNA replication starts at a specific locus called *oriC* (Figure 4).

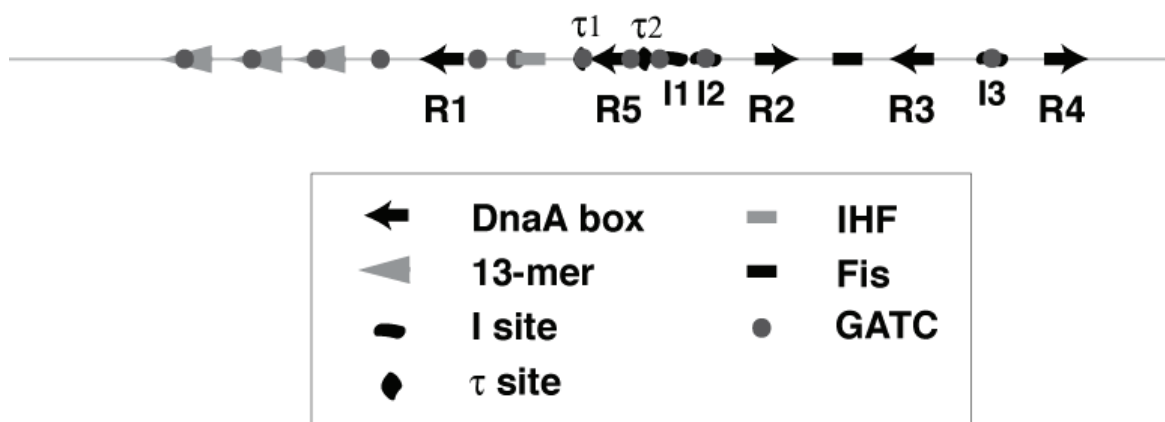


Fig. 4. Organization of DNA sequence motifs in the *E. coli* replication origin (*oriC*). Near the left border are 13-mer motifs that are unwound by DnaA complexed to ATP. Sites recognized by DnaA are the DnaA boxes (arrow), I-sites (warped oval), and  $\tau$ -sites (warped circle). Sites recognized by IHF (shaded rectangle), Fis (filled rectangle), and DNA adenine methyltransferase (shaded circle) are also indicated.

Recent reviews describe the independent mechanisms that control the frequency of initiation from this site (Nielsen and Lobner-Olesen, 2008; Katayama et al., 2010). In *Escherichia coli*, the minimal *oriC* sequence of 245 base pairs contains DNA-binding sites for many different proteins that either act directly in DNA replication, or modulate the frequency of this process (reviewed in Leonard and Grimwade, 2009). One of them is DnaA, which is the initiator of DNA replication, and has been placed in one of the clades of the AAA+ superfamily via bioinformatics analysis (Koonin, 1992; Erzberger and Berger, 2006). DnaA binds to a consensus 9 base pair sequence known as the DnaA box. There are five DnaA boxes individually named R1 through R5 within *oriC* that are similar in sequence and are recognized by DnaA (Leonard and Grimwade, 2009). In addition to these sites, DnaA complexed to ATP specifically recognizes three I-sites and  $\tau$ -sites in *oriC*, which leads to the unwinding of three AT-rich 13-mer repeats located in the left half of *oriC*. Binding sites are also present for IHF protein (integration host factor) and FIS protein (factor for inversion stimulation). As these proteins induce bends in DNA, their apparent ability to modulate the binding of DnaA to the respective sites in *oriC* may involve DNA bending. Additionally, *oriC* carries 11 GATC sequences recognized by DNA adenine methyltransferase, and sites

recognized by IciA, Rob, and SeqA. The influence of IHF, FIS, IciA, Rob and SeqA proteins on the initiation process is described in more detail in a review (Leonard and Grimwade, 2005).

At the initiation stage of DNA replication, the first step requires the binding of DnaA molecules, each complexed to ATP, to the five DnaA boxes, I- and  $\tau$ - sites of *oriC*. After binding, DnaA unwinds the duplex DNA in the AT-rich region to form an intermediate named the open complex. HU or IHF stimulates the formation of the open complex. In the next step, the replicative helicase named DnaB becomes stably bound to the separated strands of the open complex to form an intermediate named the prepriming complex. At this juncture, DnaC must be complexed to DnaB for a single DnaB hexamer to load onto each of the separated strands. DnaC protein must then dissociate from the complex in order for DnaB to be active as a helicase. Following the loading of DnaB, this helicase enlarges the unwound region of *oriC*, and then interacts with DnaG primase (Tougu and Marians, 1996). This interaction between DnaB and DnaG primase, which synthesize primer RNAs that are extended by DNA polymerase III holoenzyme during semi-conservative DNA replication, marks the transition between the process of initiation and elongation stage of DNA replication (McHenry, 2003; Corn and Berger, 2006; Langston et al., 2009). Replication fork movement that is supported by DnaB helicase and assisted by a second helicase named Rep proceeds bidirectionally around the chromosome until it reaches the terminus region (Guy et al., 2009). The two progeny DNAs then segregate near opposite poles of the cell before septum formation and cell division.

DnaC protein (27 kDa) is essential for cell viability because it is required during the initiation stage of DNA replication (reviewed in Kornberg and Baker, 1992; Davey and O'Donnell, 2003). DnaC is also required for DNA replication of the single stranded DNA of phage  $\phi$ X174, and for many plasmids (e.g. pSC101, P1, R1). DnaC additionally acts to resurrect collapsed replication forks that appear when a replication fork encounters a nick, gap, double-stranded break, or modified bases in the parental DNA (Sandler et al., 1996). This process of restarting a replication fork involves assembly of the replication restart primosome that contains PriA, PriB, PriC, DnaT, DnaB, DnaC, and Rep protein (Sandler, 2005; Gabbai and Marians, 2010). The major roles of DnaC at *oriC*, at the replication origins of the plasmids and bacteriophage described above, or in restarting collapsed replication forks is to form a complex with DnaB, which is required to escort the helicase onto the DNA, and then to depart. Since the discovery of the *dnaC* gene over 40 years ago (Carl, 1970), its ongoing study by various laboratories using a variety of approaches continue to reveal new aspects of the molecular mechanisms of DnaC in DNA replication.

Biochemical analysis combined with the X-ray crystallographic structure of the majority of *Aquifex aeolicus* DnaC (residues 43 to the C-terminal residue at position 235) reveals that DnaC protein consists of a smaller N-terminal domain that is responsible for binding to the C-terminal face of DnaB helicase, and larger ATP-binding region of 190 amino acids (Figure 2; (Ludlam et al., 2001; Galletto et al., 2003; Mott et al., 2008)). Sequence comparison of homologues of the *dnaC* gene classifies DnaC as a member of the AAA+ family of ATPases (Koonin, 1992; Davey et al., 2002; Mott et al., 2008). However unlike other AAA+ proteins, DnaC contains two additional  $\alpha$  helices named the ISM motif (Initiator/loader-Specific Motif) that directs the oligomerization of this protein into a right-handed helical filament (Mott et al., 2008). In contrast, the majority AAA+ proteins lacking these  $\alpha$  helices assemble into a closed-ring. Phylogenetic analysis of the AAA+ domain reveals that DnaC is most

closely related to DnaA, suggesting that both proteins arose from a common ancestor (Koonin, 1992). In support, the X-ray crystallographic structures of the ATPase region of DnaA and DnaC are very similar (Erzberger et al., 2006; Mott et al., 2008).

For DnaC, ATP increases its affinity for single-stranded DNA, which stimulates its ATPase activity (Davey et al., 2002; Biswas et al., 2004). Other results suggest that ATP stabilizes the interaction of DnaC with DnaB in the DnaB-DnaC complex (Wahle et al., 1989; Allen and Kornberg, 1991), which contradicts studies that support the contrary conclusion that ATP is not necessary for DnaC to form a stable complex with DnaB (Davey et al., 2002; Galletto et al., 2003; Biswas and Biswas-Fiss, 2006). As mutant DnaC proteins bearing amino acid substitutions in the Walker A box are both defective in ATP binding and apparently fail to interact with DnaB, the consequence is that these mutants cannot escort DnaB to *oriC* (Ludlam et al., 2001; Davey et al., 2002). Hence, despite the ability of DnaB by itself to bind to single-stranded DNA in vitro (LeBowitz and McMacken, 1986), DnaC is essential for DnaB to become stably bound to the unwound region of *oriC* (Kobori and Kornberg, 1982; Ludlam et al., 2001). The observation that DnaC complexed to ATP interacts with DnaA raises the possibility that both proteins act jointly in helicase loading (Mott et al., 2008). Together, these observations indicate that the ability of DnaC to bind to ATP is essential for its function in DNA replication, but the paradox about the role of ATP binding and its hydrolysis on the activity of DnaC and about the mechanism that leads to the dissociation of DnaC from DnaB have been long-standing issues.

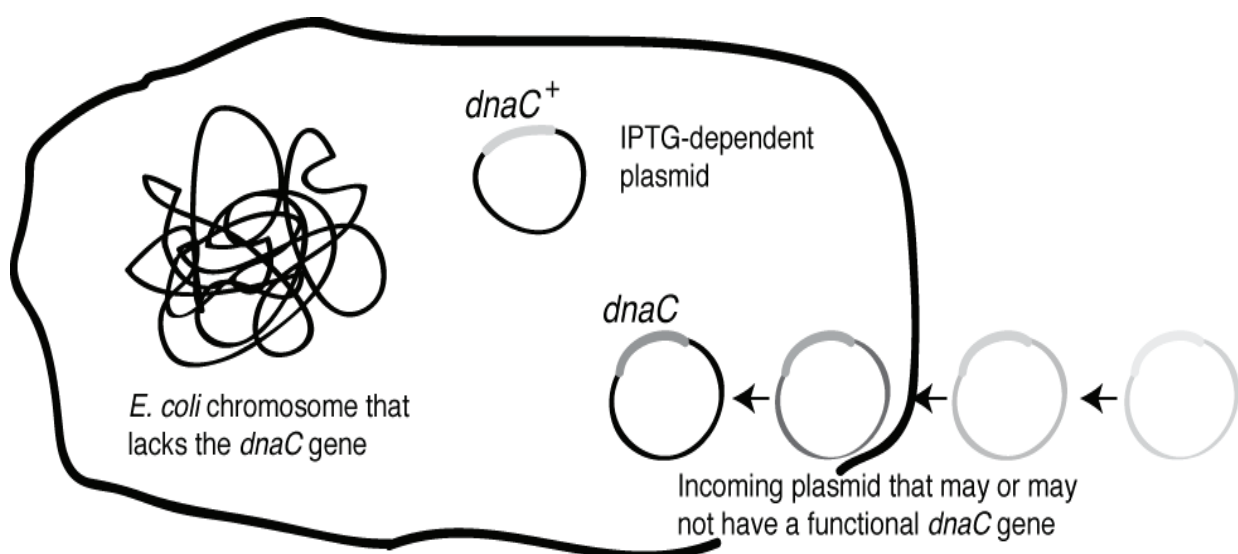


Fig. 5. Plasmid shuffle method. With an *E. coli* strain lacking the chromosomal copy of *dnaC*, a plasmid carrying the wild type *dnaC* gene residing in the bacterial cell can provide for *dnaC* function. Using a plasmid that depends on IPTG for its maintenance, the strain will not survive in the absence of IPTG. An introduced plasmid that does not depend on IPTG for its propagation can sustain viability in the absence of IPTG only if it encodes a functional *dnaC* allele.

As described above, one of the characteristics of AAA+ proteins is the presence of a conserved motif named Box VII, which carries a conserved arginine called the “arginine finger”. Structural studies of other AAA+ proteins have led to the proposal that this arginine interacts with the  $\gamma$  phosphate of ATP to promote and coordinate ATP hydrolysis with a conformational change. Recent experiments were performed to examine the role of the arginine finger of DnaC and to attempt to clarify how ATP binding and its hydrolysis by DnaC are involved in the process of initiation of DNA replication (Makowska-Grzyska and Kaguni, 2010). Part of this study relied on an *E. coli* mutant lacking the chromosomal copy of the *dnaC* gene (Hupert-Kocurek et al., 2007). Because the *dnaC* gene is essential, this deficiency of the host strain can be complemented by a plasmid encoding the *dnaC* gene that depends on IPTG (isopropyl  $\beta$ -D-1-thiogalactopyranoside) for plasmid maintenance (Figure 5). If another plasmid is introduced into the null *dnaC* mutant, it maintains viability of the strain in the absence of IPTG only if it carries a functional *dnaC* allele. In contrast, if the second plasmid carries an inactivating mutation in *dnaC*, the host strain cannot survive without IPTG. This plasmid exchange method showed that an alanine substitution for the arginine finger residue inactivated DnaC (Makowska-Grzyska and Kaguni, 2010). Biochemical experiments performed in parallel showed that this conserved arginine plays a role in the signal transduction process that involves ATP hydrolysis by DnaC that then leads to the release of DnaC from DnaB. Finally, the interaction of primase with DnaB that is coupled with primer formation is also apparently necessary for DnaC to dissociate from DnaB.

## 6. Conclusions

In summary, the combination of various experimental approaches on the study of DnaC have led to insightful experiments that expand our understanding of the role of ATP binding and its hydrolysis by DnaC during the initiation of DNA replication. Evidence suggests that ATP hydrolysis by DnaC that leads to the dissociation of DnaC from DnaB helicase is coupled with primer formation that requires an interaction between DnaG primase and DnaB. Hence, these critical steps are involved in the transition from the process of initiation to the elongation phase of DNA replication in *E. coli*.

This example on the molecular mechanism of DnaC protein is a focused study of one protein and its interaction with other required proteins during the process of initiation of DNA replication. One may consider this a form of vertical thinking. It contrasts with bioinformatics approaches that yield large sets of data for proteins based on the DNA sequences of genomes, and with microarray approaches that, for example, survey the expression of genes and their regulation at the genome level under different conditions, or identify interacting partners for a specific protein. The vast wealth of data from these global approaches provide a different perspective on understanding the functions of sets of genes or proteins and how they act in a network of biochemical pathways of the cell.

## 7. Acknowledgements

We thank members of our labs for discussions on the content and organization of this chapter. This work was supported by a grant GM090063 from the National Institutes of Health, and the Michigan Agricultural Station to JMK.

## 8. References

- Allen, G. J. and A. Kornberg (1991). Fine balance in the regulation of DnaB helicase by DnaC protein in replication in *Escherichia coli*. *J Biol Chem* 266(33): 22096-22101
- Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner and H. Mori (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2: 1-11
- Beyer, A. (1997). Sequence analysis of the AAA protein family. *Protein Sci* 6(10): 2043-58, 0961-8368 (Print), 0961-8368 (Linking)
- Biswas, S. B. and E. E. Biswas-Fiss (2006). Quantitative analysis of binding of single-stranded DNA by *Escherichia coli* DnaB helicase and the DnaB-DnaC complex. *Biochemistry* 45(38): 11505-11513
- Biswas, S. B., S. Flowers and E. E. Biswas-Fiss (2004). Quantitative analysis of nucleotide modulation of DNA binding by the DnaC protein of *Escherichia coli*. *Biochem J* 379: 553-562
- Bochner, B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 33(1): 191-205, 0168-6445 (Print), 0168-6445 (Linking)
- Carl, P. L. (1970). *Escherichia coli* mutants with temperature-sensitive synthesis of DNA. *Mol Gen Genet* 109(2): 107-122
- Carthew, R. W. and E. J. Sontheimer (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4): 642-55, 1097-4172 (Electronic), 0092-8674 (Linking)
- Christensen, N. M., K. J. Oparka and J. Tilsner (2010). Advances in imaging RNA in plants. *Trends Plant Sci* 15(4): 196-203, 1878-4372 (Electronic), 1360-1385 (Linking)
- Corn, J. E. and J. M. Berger (2006). Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Res* 34(15): 4082-4088
- Davey, M. J., L. Fang, P. McInerney, R. E. Georgescu and M. O'Donnell (2002). The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J* 21(12): 3148-3159
- Davey, M. J., D. Jeruzalmi, J. Kuriyan and M. O'Donnell (2002). Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol* 3(11): 826-835
- Davey, M. J. and M. O'Donnell (2003). Replicative helicase loaders: ring breakers and ring makers. *Curr Biol* 13(15): R594-596
- DePamphilis, M. L. and S. D. Bell (2010). *Genome duplication*, Garland Science/Taylor & Francis Group, 9780415442060, London
- Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411(6836): 494-8, 0028-0836 (Print), 0028-0836 (Linking)
- Erzberger, J. P. and J. M. Berger (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* 35: 93-114
- Erzberger, J. P., M. L. Mott and J. M. Berger (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* 13(8): 676-683

- Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669): 806-11, 0028-0836 (Print), 0028-0836 (Linking)
- Fischer, S. E. (2010). Small RNA-mediated gene silencing pathways in *C. elegans*. *Int J Biochem Cell Biol* 42(8): 1306-15, 1878-5875 (Electronic), 1357-2725 (Linking)
- Gabbai, C. B. and K. J. Mariani (2010). Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival. *DNA Repair (Amst)* 9(3): 202-209, 1568-7856 (Electronic), 1568-7856 (Linking)
- Galletto, R., M. J. Jezewska and W. Bujalowski (2003). Interactions of the *Escherichia coli* DnaB Helicase Hexamer with the Replication Factor the DnaC Protein. Effect of Nucleotide Cofactors and the ssDNA on Protein-Protein Interactions and the Topology of the Complex. *J Mol Biol* 329(3): 441-465
- Ghildiyal, M. and P. D. Zamore (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10(2): 94-108, 1471-0064 (Electronic), 1471-0056 (Linking)
- Guy, C. P., J. Atkinson, M. K. Gupta, A. A. Mahdi, E. J. Gwynn, C. J. Rudolph, P. B. Moon, I. C. van Knippenberg, C. J. Cadman, M. S. Dillingham, R. G. Lloyd and P. McGlynn (2009). Rep provides a second motor at the replisome to promote duplication of protein-bound DNA. *Mol Cell* 36(4): 654-666, 1097-4164 (Electronic), 1097-2765 (Linking)
- Hamilton, A. J. and D. C. Baulcombe (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286(5441): 950-952, 0036-8075 (Print), 0036-8075 (Linking)
- Hanson, P. I. and S. W. Whiteheart (2005). AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* 6(7): 519-529
- Hupert-Kocurek, K., J. M. Sage, M. Makowska-Grzyska and J. M. Kaguni (2007). Genetic method to analyze essential genes of *Escherichia coli*. *Appl Environ Microbiol* 73(21): 7075-7082
- Iyer, L. M., D. D. Leipe, E. V. Koonin and L. Aravind (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146(1-2): 11-31, 1047-8477 (Print), 1047-8477 (Linking)
- Katayama, T., S. Ozaki, K. Keyamura and K. Fujimitsu (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat Rev Microbiol* 8(3): 163-170, 1740-1534 (Electronic), 1740-1526 (Linking)
- Ketting, R. F., T. H. Haverkamp, H. G. van Luenen and R. H. Plasterk (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99(2): 133-141, 0092-8674 (Print), 0092-8674 (Linking)
- Kobori, J. A. and A. Kornberg (1982). The *Escherichia coli dnaC* gene product. II. Purification, physical properties, and role in replication. *J Biol Chem* 257(22): 13763-13769
- Koonin, E. V. (1992). DnaC protein contains a modified ATP-binding motif and belongs to a novel family of ATPases including also DnaA. *Nucleic Acids Res* 20(8): 1997

- Kornberg, A. and T. A. Baker (1992). DNA Replication Second Edition, W.H. Freeman and Company, 9781891389443, New York
- Krol, J., I. Loedige and W. Filipowicz (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11(9): 597-610, 1471-0064 (Electronic), 1471-0056 (Linking)
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen (2001).

- Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921, 0028-0836 (Print), 0028-0836 (Linking)
- Langston, L. D., C. Indiani and M. O'Donnell (2009). Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle* 8(17): 2686-2691, 1551-4005 (Electronic), 1551-4005 (Linking)
- LeBowitz, J. H. and R. McMacken (1986). The *Escherichia coli* dnaB replication protein is a DNA helicase. *J Biol Chem* 261(10): 4738-48,
- Lee, D. G. and S. P. Bell (2000). ATPase switches controlling DNA replication initiation. *Curr Opin Cell Biol* 12(3): 280-285, 0955-0674 (Print), 0955-0674 (Linking)
- Leonard, A. C. and J. E. Grimwade (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Mol Microbiol* 55(4): 978-985
- Leonard, A. C. and J. E. Grimwade (2009). Initiating chromosome replication in *E. coli*: it makes sense to recycle. *Genes Dev* 23(10): 1145-1150,
- Liu, J., C. L. Smith, D. DeRyckere, K. DeAngelis, G. S. Martin and J. M. Berger (2000). Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell* 6(3): 637-648
- Liu, Q., T. A. Rand, S. Kalidas, F. Du, H. E. Kim, D. P. Smith and X. Wang (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 301(5641): 1921-1925, 1095-9203 (Electronic), 0036-8075 (Linking)
- Ludlam, A. V., M. W. McNatt, K. M. Carr and J. M. Kaguni (2001). Essential amino acids of *Escherichia coli* DnaC protein in an N-terminal domain interact with DnaB helicase. *J Biol Chem* 276(29): 27345-27353
- Lupas, A. N. and J. Martin (2002). AAA proteins. *Curr Opin Struct Biol* 12(6): 746-53, 0959-440X (Print), 0959-440X (Linking)
- Lupski, J. R., J. G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D. C. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D. A. Wheeler, A. L. McGuire, F. Zhang, P. Stankiewicz, J. J. Halperin, C. Yang, C. Gehman, D. Guo, R. K. Irikat, W. Tom, N. J. Fantin, D. M. Muzny and R. A. Gibbs (2010). Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362(13): 1181-91, 1533-4406 (Electronic), 0028-4793 (Linking)
- Makowska-Grzyska, M. and J. M. Kaguni (2010). Primase Directs the Release of DnaC from DnaB. *Mol Cell* 37(1): 90-101
- Maxam, A. M. and W. Gilbert (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74(2): 560-4, 0027-8424 (Print), 0027-8424 (Linking)
- McHenry, C. S. (2003). Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences. *Mol Microbiol* 49(5): 1157-1165
- Moerman, D. G. and R. J. Barstead (2008). Towards a mutation in every gene in *Caenorhabditis elegans*. *Brief Funct Genomic Proteomic* 7(3): 195-204, 1477-4062 (Electronic), 1473-9550 (Linking)
- Mott, M. L., J. P. Erzberger, M. M. Coons and J. M. Berger (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* 135(4): 623-634



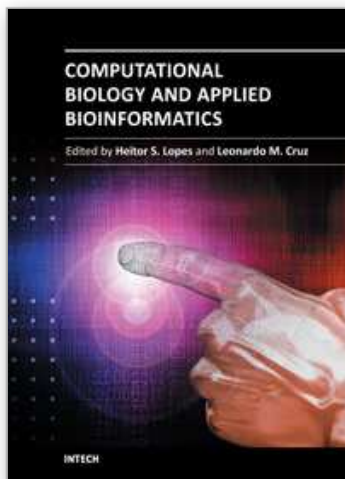
- Neuwald, A. F., L. Aravind, J. L. Spouge and E. V. Koonin (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 9(1): 27-43
- Ng, P. C. and E. F. Kirkness (2010). Whole genome sequencing. *Methods Mol Biol* 628: 215-226, 1940-6029 (Electronic), 1064-3745 (Linking)
- Nielsen, O. and A. Lobner-Olesen (2008). Once in a lifetime: strategies for preventing re-replication in prokaryotic and eukaryotic cells. *EMBO Rep* 9(2): 151-156
- Ogura, T., S. W. Whiteheart and A. J. Wilkinson (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J Struct Biol* 146(1-2): 106-112
- Reznikoff, W. S. and K. M. Winterberg (2008). Transposon-based strategies for the identification of essential bacterial genes. *Methods Mol Biol* 416: 13-26, 1064-3745 (Print), 1064-3745 (Linking)
- Sandler, S. J. (2005). Requirements for replication restart proteins during constitutive stable DNA replication in *Escherichia coli* K-12. *Genetics* 169(4): 1799-1806
- Sandler, S. J., H. S. Samra and A. J. Clark (1996). Differential suppression of *priA2::kan* phenotypes in *Escherichia coli* K-12 by mutations in *priA*, *lexA*, and *dnaC*. *Genetics* 143(1): 5-13
- Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, C. A. Fiddes, C. A. Hutchison, P. M. Slocombe and M. Smith (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265(5596): 687-95, 0028-0836 (Print), 0028-0836 (Linking)
- Sanger, F., S. Nicklen and A. R. Coulson (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74(12): 5463-7, 0027-8424 (Print), 0027-8424 (Linking)
- Sutcliffe, J. G. (1979). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harb Symp Quant Biol* 43 Pt 1: 77-90, 0091-7451 (Print), 0091-7451 (Linking)
- Swaffield, J. C. and M. D. Purugganan (1997). The evolution of the conserved ATPase domain (CAD): reconstructing the history of an ancient protein module. *J Mol Evol* 45(5): 549-563, 0022-2844 (Print), 0022-2844 (Linking)
- Tabara, H., M. Sarkissian, W. G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire and C. C. Mello (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99(2): 123-32, 0092-8674 (Print), 0092-8674 (Linking)
- Tougu, K. and K. J. Marians (1996). The interaction between helicase and primase sets the replication fork clock. *J Biol Chem* 271(35): 21398-405,
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R.

- Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu (2001). The sequence of the human genome. *Science* 291(5507): 1304-1351, 0036-8075 (Print), 0036-8075 (Linking)
- Vidan, S. and M. Snyder (2001). Large-scale mutagenesis: yeast genetics in the genome era. *Curr Opin Biotechnol* 12(1): 28-34, 0958-1669 (Print), 0958-1669 (Linking)
- Wahle, E., R. S. Lasken and A. Kornberg (1989). The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. *J Biol Chem* 264(5): 2463-2468
- Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, N. Liebundguth, D. J. Lockhart, A. Lucau-Danila, M. Lussier, N. M'Rabet, P. Menard, M. Mittmann, C. Pai, C. Rebischung, J. L. Revuelta, L. Riles, C. J. Roberts, P. Ross-MacDonald, B. Scherens, M. Snyder, S. Sookhai-Mahadeo, R. K. Storms, S. Veronneau, M. Voet, G.

Volckaert, T. R. Ward, R. Wysocki, G. S. Yen, K. Yu, K. Zimmermann, P. Philippsen, M. Johnston and R. W. Davis (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285(5429): 901-906, 0036-8075 (Print), 0036-8075 (Linking)

IntechOpen

IntechOpen



## **Computational Biology and Applied Bioinformatics**

Edited by Prof. Heitor Lopes

ISBN 978-953-307-629-4

Hard cover, 442 pages

**Publisher** InTech

**Published online** 02, September, 2011

**Published in print edition** September, 2011

Nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics. It is not different with biology, that has seen an unpredictable growth in recent decades, with the rise of a new discipline, bioinformatics, bringing together molecular biology, biotechnology and information technology. More recently, the development of high throughput techniques, such as microarray, mass spectrometry and DNA sequencing, has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. On the other hand, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. The development of bioinformatics has allowed now the emergence of systems biology, the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of a living being. This book presents some theoretical issues, reviews, and a variety of bioinformatics applications. For better understanding, the chapters were grouped in two parts. In Part I, the chapters are more oriented towards literature review and theoretical issues. Part II consists of application-oriented chapters that report case studies in which a specific biological problem is treated with bioinformatics tools.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Katarzyna Hupert-Kocurek and Jon M. Kaguni (2011). Understanding Protein Function - The Disparity Between Bioinformatics and Molecular Methods, Computational Biology and Applied Bioinformatics, Prof. Heitor Lopes (Ed.), ISBN: 978-953-307-629-4, InTech, Available from:  
<http://www.intechopen.com/books/computational-biology-and-applied-bioinformatics/understanding-protein-function-the-disparity-between-bioinformatics-and-molecular-methods>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

[www.intechopen.com](http://www.intechopen.com)

[www.intechopen.com](http://www.intechopen.com)

IntechOpen

IntechOpen

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen