

Revisiting the Krogh Principle in the post-genome era: *Caenorhabditis elegans* as a model system for integrative physiology research

Kevin Strange

Departments of Anesthesiology, Molecular Physiology and Biophysics, and Pharmacology, Vanderbilt University
 Medical Center, Nashville, TN 37232, USA

e-mail: kevin.strange@vanderbilt.edu

Accepted 12 December 2006

Summary

Molecular biology drove a powerful reductionist or 'molecule-centric' approach to biological research in the last half of the 20th century. Reductionism is the attempt to explain complex phenomena by defining the functional properties of the individual components that comprise multi-component systems. Systems biology has emerged in the post-genome era as the successor to reductionism. In my opinion, systems biology and physiology are synonymous. Both disciplines seek to understand multi-component processes or 'systems' and the underlying pathways of information flow from an organism's genes up through increasingly complex levels of organization.

The physiologist and Nobel laureate August Krogh believed that there is an ideal organism in which almost every physiological problem could be studied most readily (the 'Krogh Principle'). If an investigator's goal were to define a physiological process from the level of genes to the whole animal, the optimal model organism for him/her to utilize would be one that is genetically and molecularly tractable. In other words, an organism in which forward and reverse genetic analyses could be carried out readily,

rapidly and economically. Non-mammalian model organisms such as *Escherichia coli*, *Saccharomyces*, *Caenorhabditis elegans*, *Drosophila*, zebrafish and the plant *Arabidopsis* are cornerstones of systems biology research.

The nematode *C. elegans* provides a particularly striking example of the experimental utility of non-mammalian model organisms. The aim of this paper is to illustrate how genetic, functional genomic, molecular and physiological methods can be combined in *C. elegans* to develop a systems biological understanding of fundamental physiological processes common to all animals. I present examples of the experimental tools available for the study of *C. elegans* and discuss how we have used them to gain new insights into osmotic stress signaling in animal cells.

Glossary available online at
<http://jeb.biologists.org/cgi/content/full/210/9/1622/DC1>

Key words: *C. elegans*, Krogh Principle, genomics, osmotic stress.

Introduction

Physiology research in the last half of the 20th century was dominated by a powerful reductionist or 'molecule-centric' approach. Reductionism attempts to explain complex phenomena by defining the functional properties of the individual components that comprise multi-component systems. Genome sequencing has ushered in the end of what Bloom termed 'naïve reductionism' (Bloom, 2001). Reductionist methods will continue to be an essential element of all biological research efforts, but 'naïve reductionism', the belief that reductionism alone can lead to a complete understanding of living organisms, is not tenable. Organisms are clearly much more than the sum of their parts and the behavior of complex physiological processes cannot be

understood simply by knowing how the parts work in isolation.

The post-genome sequencing era can rightfully be thought of as the era of integrative biology or, to use the more current catchphrase, systems biology (Ideker et al., 2001; Kitano, 2002a; Kitano, 2002b; Pennisi, 2003). Integrative/systems biology seeks to understand and predict the behavior or 'emergent' properties of complex, multi-component biological processes. An integrative/systems level molecular characterization of a biological process addresses three main questions. (1) What are the parts of the system (i.e. the genes and the proteins they encode)? (2) How do the parts work? (3) And most importantly, how do the parts work together to accomplish a task?

The Krogh Principle

August Krogh is widely viewed as a leading figure in the field of integrative physiology. Over the course of his career, Krogh made numerous fundamental contributions to our understanding of gas exchange and respiration, capillary blood flow, ion and water exchange, and exercise physiology and metabolism. In 1920, Krogh won the Nobel Prize in Physiology or Medicine for his work on capillary blood flow and oxygen utilization during muscular work.

Krogh's seminal contributions to physiology reflect his intuition for choosing important problems and for developing the right tools and experimental strategies to address those problems. His scientific intuition is also reflected in his choice of the right experimental model. In this regard, Krogh was very much a comparative physiologist and is famous for stating that, "for many problems there is an animal in which it can be most conveniently studied", a statement that has subsequently become known as the Krogh Principle (Krebs, 1975; Krogh, 1929).

Integrative physiology and genetic model organisms

Genome sequencing coupled with stunning technological advances has made it feasible to take integrative physiology research to the most basic molecular level and define physiological processes beginning with their underlying genes and protein networks. If an integrative physiologist's goal was to define the genes and integrated genetic pathways underlying a physiological process, in what animal model could the problem be best addressed? The ideal organism in which the problem could be 'most conveniently studied' would have a fully sequenced and well-annotated genome. Importantly, the model system should be genetically tractable. Mutagenesis and forward genetic analysis allows one to identify genes relevant to a process of interest in a completely unbiased manner and allows assembly of those genes into functional pathways. The model should also be molecularly tractable. In other words, the model should allow straightforward and hopefully economical manipulation of gene expression through reverse genetic strategies and transgenesis.

Genetically tractable, non-mammalian model organisms such as *Escherichia coli*, *Saccharomyces*, *Caenorhabditis elegans*, *Drosophila*, zebrafish and the plant *Arabidopsis* are cornerstones of modern biomedical research. In the post-genome era, these organisms have been likened to the Rosetta Stone (Ideker et al., 2001), which provided modern scholars the tools needed to decipher Egyptian hieroglyphics. Similarly, genetic model organisms provide powerful tools that allow genome sequence to be deciphered. For the integrative physiologist then, a genetically tractable model organism would be an essential component of any research effort aimed at developing a genetic understanding of a physiological process.

My own research interests are focused broadly on the integrative physiology of ion and water homeostasis, particularly osmosensitive ion channels, epithelial transport and signaling mechanisms, and the cellular osmotic stress

response, and I have studied these problems in a variety of models including intertidal bivalves, saltwater mosquito larvae, mammalian kidney tubules and cells in the mammalian central nervous system including astrocytes, neurons and choroid plexus cells. In late 1998, I had grown terminally frustrated over our inability to develop an integrated molecular understanding of these problems and began searching for new experimental models. The Krogh Principle dictated that we utilize a genetically tractable organism for our studies. As an animal physiologist, the organism that interested me most was the nematode *C. elegans*. Hermaphrodite genetics were certainly easier to understand and utilize experimentally. In addition, the animal's relative 'simplicity' was appealing. *C. elegans* is complex enough to be interesting, but its simple body plan and limited cell number make it experimentally more tractable than fruit flies and fish. The laboratory culture of worms is also very straightforward and was something that we could set up quickly with a minimum of cost.

In the following sections, I provide a brief summary of *C. elegans* biology and discuss its experimental attributes. The last section is a brief overview of our recent work and a description of how we have exploited the worm to address a physiological problem of broad relevance to all animals.

C. elegans biology

Natural history and life cycle

Caenorhabditis elegans is a free-living nematode about 1 mm long and is typically found inhabiting surface soil and decaying vegetable matter. The life strategy of *C. elegans* is well adapted for survival in soil environments where food and water availability, temperature, populations of predators and many other variables can change constantly and dramatically. It is a voracious feeder and outgrows its competitors by producing large numbers of offspring and rapidly depleting local food resources.

Adult *C. elegans* are predominantly hermaphroditic with males making up approximately 0.1% of wild-type populations. Self-fertilized hermaphrodites produce about 300 offspring whereas male-fertilized hermaphrodites can produce over 1000 progeny. Postembryonic development occurs in four larval stages (L1–L4) and adult worms survive about 2–3 weeks under optimal laboratory conditions.

When food supply is limited, dauer larvae form after the second larval molt. Dauer larvae do not feed and have structural, metabolic and behavioral adaptations that increase life span up to 10 times and aid in the dispersal of the animal to new habitats. Once food becomes available, dauer larvae feed and continue development to the adult stage (Riddle and Albert, 1997).

Laboratory culture

Culture of *C. elegans* in the laboratory is simple and relatively inexpensive (Lewis and Fleming, 1995). Animals are typically grown in Petri dishes on agar seeded with a lawn of *E. coli* as a food source. *C. elegans* can also be grown in mass

quantities using liquid culture strategies and fermentor-like devices. Worm stocks are stored frozen in liquid nitrogen indefinitely with good viability, which greatly simplifies culture strategies and reduces costs associated with handling and maintaining wild type and mutant worm strains.

Anatomy

Like all nematodes, *C. elegans* has an unsegmented, cylindrical body that tapers at both ends. The body wall consists of a tough collagenous cuticle underlain by hypodermis, muscles and nerves. A fluid-filled body cavity or pseudocoel separates the body wall from internal organs. Body shape is maintained by hydrostatic pressure in the pseudocoel.

Newly hatched L1 larvae have 558 cells. Additional divisions of somatic blast cells occur during the four larval stages eventually giving rise to 959 somatic cells in mature adult hermaphrodites and 1031 in adult males. The lineage of somatic cells in *C. elegans* is largely invariant. This invariance, combined with the ability to visualize by differential interference contrast microscopy cell division and development in living embryos, larvae and adult animals, has made it possible to describe the fate map or cell lineage of the worm (Sulston et al., 1983; Sulston and Horvitz, 1977).

Despite the small cell number, *C. elegans* exhibits a striking degree of differentiation. Many physiological functions found in mammals have nematode analogs. This high degree of complexity and small total cell number provides a remarkably tractable experimental system for studies of differentiation, cell biology and cell physiology. A detailed description of worm anatomy can be found online at the Center for *C. elegans* Anatomy (<http://www.aecom.yu.edu/wormem/>).

C. elegans has a well developed musculature and nervous system and has proved to be an invaluable model system for the study of excitable cell physiology. The worm possesses both striated and non-striated muscles. Striated body wall muscles are the most numerous muscle cell type and are responsible for locomotion. Non-striated muscles are associated with the pharynx, intestine, anus and gonad, and mediate pharyngeal pumping, defecation, ovulation and fertilization, and egg laying.

The nervous system of adult hermaphrodites contains 302 neurons and 56 glial and support cells. Males have 381 neurons and 92 glial and support cells. White et al. (White et al., 1986) have reconstructed and mapped the connectivity of the entire hermaphrodite nervous system using serial electron microscopy. Most of the differences between the male and hermaphrodite nervous system are found in the male tail, which plays an important role in mating. An important feature of the *C. elegans* nervous system is that only three neurons, which control pharyngeal pumping required for feeding and fluid excretion by the excretory cell, are essential for survival under laboratory conditions. The nonessential nature of most neurons for viability provides an enormous advantage for mutagenesis studies of nervous system function.

The worm 'kidney' consists of three cell types, the excretory cell, the duct cell and the pore cell (Nelson et al., 1983).

Destruction of any of these cells by laser ablation causes the animal to swell with fluid and die (Nelson and Riddle, 1984). The excretory cell is a large, H-shaped cell that sends out processes both anteriorly and posteriorly from the cell body. A fluid-filled excretory canal is surrounded by the cell cytoplasm. The basal pole of the cell faces the pseudocoel while the apical membrane faces the excretory canal lumen. Gap junctions connect the excretory cell to the hypodermis, an epithelium that lies just below the cuticle. An excretory duct connects the excretory canal to the outside surface of the worm and is formed by the duct and pore cells.

The digestive tract of *C. elegans* consists of a pharynx, intestine and rectum. *C. elegans* is a filter feeder and the pharynx is a muscular organ that pumps food into the pharyngeal lumen, grinds it up and then moves it into the intestine. The pharynx is formed from muscle cells, neurons, epithelial cells and gland cells (Albertson and Thomson, 1976). Twenty epithelial cells with extensive apical microvilli form the main body of the intestine (Leung et al., 1999). Intestinal epithelial cells secrete digestive enzymes and absorb nutrients.

As noted above, sexual reproduction in *C. elegans* occurs by self-fertilization of hermaphrodites or fertilization of hermaphrodites by males. The gonad of adult hermaphrodites consists of two identical U-shaped tubes connected *via* spermatheca to a common uterus. Sperm are formed during the fourth larval stage and stored in the spermatheca. In adults, germ cells develop into oocytes and are ovulated into the spermatheca for fertilization. The male gonad consists of the testis, seminal vesicle and vas deferens.

Forward genetic screening

The development of *C. elegans* as an experimental system was driven largely by the relative ease of performing forward genetic screens for identification of the complement of genes responsible for observable phenotypes. The utility and power of genetic screening depends on the ability to assay a phenotype of interest. For a detailed discussion of screening assays in *C. elegans*, the reader is referred to a recent review (Jorgensen and Mango, 2002) as well as several chapters in WormBook (<http://www.wormbook.org>), an online review of *C. elegans* biology.

Once a screening assay is developed, animals are mutagenized, typically by the alkylating agent ethyl methane-sulphonate (EMS). Mutant animals are then isolated and the mutated gene identified by mapping, rescue and cloning strategies. The reproductive characteristics of *C. elegans* greatly facilitate the isolation and maintenance of mutant strains. Self-fertilization in hermaphrodites allows homozygous animals to breed true and is especially useful if mutant worms are paralyzed or uncoordinated since reproduction does not require movement in order to find and mate with a male. Mating with males, however, is essential for moving mutations between strains.

Mutant animals can be further mutagenized to suppress or enhance the original phenotype. Suppressor or enhancer

mutations may reside in genes distinct from the one mutated in the original screen. These extragenic mutations imply that the suppressor and enhancer genes interact with the first mutated gene. Genetic interactions indicate that gene products function in a common process.

Reverse genetics

One of the truly extraordinary experimental advantages of *C. elegans* is the relative ease by which gene expression can be silenced or knocked down using double stranded RNA (dsRNA)-mediated gene interference (RNAi) (Sen and Blau, 2006). RNAi is induced in worms by injecting them with dsRNA (Fire et al., 1998), by soaking them in dsRNA solutions (Tabara et al., 1998) or by feeding them bacteria producing dsRNA (Kamath et al., 2000; Timmons et al., 2001; Timmons and Fire, 1998). When worms are fed dsRNA-producing bacteria or soaked in dsRNA solutions, the dsRNA is absorbed across the intestinal epithelium and then spreads systemically to the animal's somatic cells and germline. In cultured *C. elegans* cells, RNAi is triggered simply by adding dsRNA to the culture medium (Christensen et al., 2002).

Kamath et al. generated a reusable RNAi library (available from Geneservice Ltd, Cambridge, UK) consisting of ~16 000 bacterial strains, each of which expresses a unique dsRNA (Kamath et al., 2003). A second RNAi library generated by Vidal and coworkers [(Rual et al., 2004); available from Open Biosystems, Huntsville, AL, USA] contains ~11 800 dsRNA-producing bacterial strains. Together, these two libraries provide RNAi bacterial clones to ~90% of the genes in the worm genome. A number of genome-wide RNAi screens have been carried out in *C. elegans* and have successfully identified genes involved in fundamental biological processes including fat metabolism (Ashrafi et al., 2003), ageing (Lee et al., 2003; Murphy et al., 2003), early embryonic development (Zipperlen et al., 2001), osmotic stress resistance (Lamitina and Strange, 2004; Lamitina et al., 2006) and prevention of protein aggregation (Nollen et al., 2004).

Gene knockout or inactivation is another important reverse genetic strategy. Targeted gene knockout by homologous recombination using microparticle bombardment methods (Berezikov et al., 2004) or DNA microinjection into meiotic oocyte nuclei (Broverman et al., 1993) has been reported in *C. elegans*, but has not been widely used as an experimental tool. Instead, the relative ease of culturing *C. elegans* in large numbers and the ability to store worms frozen has led to the development of so-called 'target-selected gene inactivation methods'. This approach involves inducing random deletion mutations in a population of worms using either chemical mutagens or transposons (e.g. Jansen et al., 1997; Williams et al., 2005). Several large-scale efforts to produce strains possessing deletion mutations in all identified worm genes are underway (e.g. <http://elegans.bcgsc.bc.ca/knockout.shtml>; <http://shigen.lab.nig.ac.jp/c.elegans/index.jsp>; <http://elegans.imbb.forth.gr/nemagenetag/>). Once strains are created, they are made freely available to the research community.

Creation of transgenic worms

DNA transformation in *C. elegans* is relatively straightforward (Fire, 1986; Mello et al., 1991; Stinchcomb et al., 1985). Briefly, transforming DNA is microinjected into the distal end of the hermaphrodite gonad. Heritable DNA transformation occurs by extrachromosomal transformation, nonhomologous integration or homologous integration. Spontaneous homologous integration is extremely rare. Formation of multicopy extrachromosomal arrays is the most frequent way in which transforming DNA is inherited. Transformation by extrachromosomal arrays is often transient. Integration of transgenes and generation of stable transgenic lines is commonly carried out by gamma irradiation of transformed worms (Mello and Fire, 1995). Microparticle bombardment can also be used to create integrated transgenic lines in *C. elegans* with a frequency of 9–35% relative to the number of bombardments performed (Praitis et al., 2001).

Tools for cell physiology

C. elegans is exceptionally well-suited for quantitative, *in vivo* microscopy (Hall et al., 2006). The embryo eggshell and cuticle of larvae and adults are transparent, making it possible to observe and quantitate cell biological events and physiological processes, including Ca²⁺ signaling (Schafer, 2005) and intracellular pH regulation (Nehrke, 2003), using brightfield and fluorescence microscopy. Electron microscopy methods for *C. elegans* are well developed (Hall et al., 2006).

A powerful way to assess the physiological role of a specific nematode cell type is to destroy the cell and characterize the effect on developmental events and whole animal phenotype. Laser ablation or microsurgery has been used extensively to identify cell function and cell–cell developmental interactions in *C. elegans* (Bargmann and Avery, 1995). It is also possible to genetically target cells for killing using transgenic methods (e.g. Harbinder et al., 1997; Maricq et al., 1995).

In vivo electrophysiology in *C. elegans* is technically demanding due to the small size of the animal. However, several elegant experimental strategies have been developed that allow *in vivo* patch clamp studies of neurons and muscle cells (e.g. Brockie et al., 2001; Goodman et al., 1998; Lockery and Goodman, 1998; Richmond et al., 1999; Richmond and Jorgensen, 1999).

The *C. elegans* pharynx has been utilized as a model system for identifying the genetic basis of ion channel and excitable cell function. Pharynx action potentials have been characterized using extracellular recording techniques (Raizen and Avery, 1994) and an isolated pharynx preparation that allows impalements with glass microelectrodes (e.g. Davis et al., 1999; Franks et al., 2002). Isolated preparations of developing embryo cells (Christensen and Strange, 2001) and oocytes (Rutledge et al., 2001) have also been used study ion channel activity and regulation.

Until recently, the culture of differentiated *C. elegans* cells was thought to be technically infeasible. However, methods that allow the robust, large-scale culture of *C. elegans* embryonic cells have now been described (Christensen et al.,

2002). Isolated embryonic cells differentiate within 24 h into the various cell types that form the newly hatched L1 larva. Cultured somatic cells have been particularly useful for electrophysiological studies of ion channel function (e.g. Christensen et al., 2002; Estevez et al., 2003; Yuan et al., 2003). Fluorescence-activated and magnetic-activated cell sorting can be used to enrich cell types of interest, in turn allowing cell-specific biochemical, molecular, DNA microarray and proteomic studies (e.g. Cinar et al., 2005; Colosimo et al., 2004; Fox et al., 2005).

Functional genomics

Developing a molecular level understanding of a physiological process requires identification of the genes, and the proteins they encode (i.e. the 'parts'), that work together to give rise to that process. Functional genomics, which utilizes large-scale and high-throughput methodologies to define and analyze gene function at a global level (Segal and Kim, 2003; Yanai, 2003), is therefore an important component of molecular integrative/systems biology research. Numerous functional genomics studies including genome-wide microarray (e.g. Shen et al., 2005; Viswanathan et al., 2005) and RNAi screens (discussed above) have been carried out in *C. elegans*. In addition, a genome-wide protein-protein interaction map is being developed for the worm (Li et al., 2004). Integration of these large-scale datasets with functional studies can provide important and novel insights into physiological processes (e.g. Boulton et al., 2002; Walhout et al., 2002; Gunsalus et al., 2005; Zhong and Sternberg, 2006).

Reagents and online resources

The 'worm community' is well known for its open sharing of data and reagents. Numerous reagents including cosmid, YAC and EST clones are freely available from public resources. Literally thousands of mutant and transgenic worm strains are maintained and available at the *Caenorhabditis* Genetics Center (<http://www.cbs.umn.edu/CGC/>). In addition, an extraordinary wealth of data on *C. elegans* is available online. Indeed, the worm community was an early pioneer in the use of the Internet for electronic data sharing. WormBase (<http://www.wormbase.org/>) is a particularly noteworthy database. It provides an exhaustive catalog of worm biology including identification of all known and predicted worm genes. Gene descriptions include genome location, mutant and RNAi phenotypes, expression patterns, microarray data, gene ontology, mutant alleles and BLAST matches (Schwarz et al., 2006). WormBook (<http://www.wormbook.org/>) is an extensive online collection of chapters describing *C. elegans* biology and methodology.

Using *C. elegans* as a model system for integrative physiology research

We have utilized *C. elegans* extensively to study Cl⁻ anion channel physiology, epithelial cell oscillatory Ca²⁺ signaling and osmotic homeostasis. For the purposes of this article, I will

provide a brief overview of our recent work on cellular osmotic stress physiology. This work illustrates use of many of the tools described above.

Fig. 1 is a cartoon showing the response of animal cells to hypertonic stress. Exposure of animal cells to hypertonic media causes rapid water loss and cell shrinkage. Most cells respond to shrinkage by activation of regulatory volume increase (RVI) salt uptake mediated by Na⁺/K⁺/2Cl⁻ cotransporters or Na⁺/H⁺ and Cl⁻/HCO₃⁻ exchange mechanisms. Osmotically obliged water follows salt uptake and cell volume returns to its original value (Lang et al., 1998).

The net effect of cell shrinkage and subsequent RVI is an increase in intracellular ionic strength. Inorganic ions are so-called 'perturbing' solutes and can cause protein denaturation and other forms of cell damage when they are present in elevated concentrations. The second phase of the hypertonic stress response then is the replacement of inorganic ions by organic osmolytes. Organic osmolytes are 'compatible' or 'non-perturbing' solutes such as glycerol, sorbitol, taurine, proline and betaine. These solutes have unique biophysical and biochemical properties that allow cells to accumulate them to high levels or to withstand large shifts in their concentration without deleterious effects on cellular structure and function. Accumulation of organic osmolytes is mediated either by energy-dependent transport from the external medium or by changes in the rates of osmolyte synthesis and degradation (Yancey, 2005). Hypertonic stress typically increases the expression of both organic osmolyte transporters and key enzymes involved in their synthesis (Jeon et al., 2006). The third component of the hypertonic stress response is the repair of molecular damage, including DNA breaks and protein denaturation, induced by the initial cell shrinkage and elevation ionic strength (e.g. Sheen et al., 2006).

While the effector mechanisms that mediate RVI and organic osmolyte accumulation in animal cells are well described, little is known about the molecular basis of the signals and signaling mechanisms that activate these pathways (Fig. 2). *C. elegans* seemed to be an ideal model system in which to define how animal cells detect osmotic stress. As noted earlier, worms inhabit decaying vegetable matter such as forest floor leaf litter. Soil environments are osmotically unstable and worms are exposed to constant osmotic challenges. In the laboratory, *C. elegans* readily survives and adapts to growth media made hypertonic by addition of up to 500 mmol l⁻¹ NaCl (Lamitina et al., 2004). Worms lose water rapidly and shrink during hypertonic stress, but then regain their original volume within a few tens of minutes. Over a period of several hours, worms accumulate the organic osmolyte glycerol. Glycerol accumulation is mediated by de novo synthesis. Glycerol 3-phosphate dehydrogenase (GPDH) catalyzes the rate-limiting step in glycerol synthesis. Two genes, *gpdh-1* and *gpdh-2*, encode this enzyme in *C. elegans*. Northern analyses demonstrated that *gpdh-1* is transcriptionally upregulated in response to hypertonic stress. Extensive genome-wide microarray studies (T. Lamitina and K.S., unpublished observations), revealed that *gpdh-1* shows a sustained and

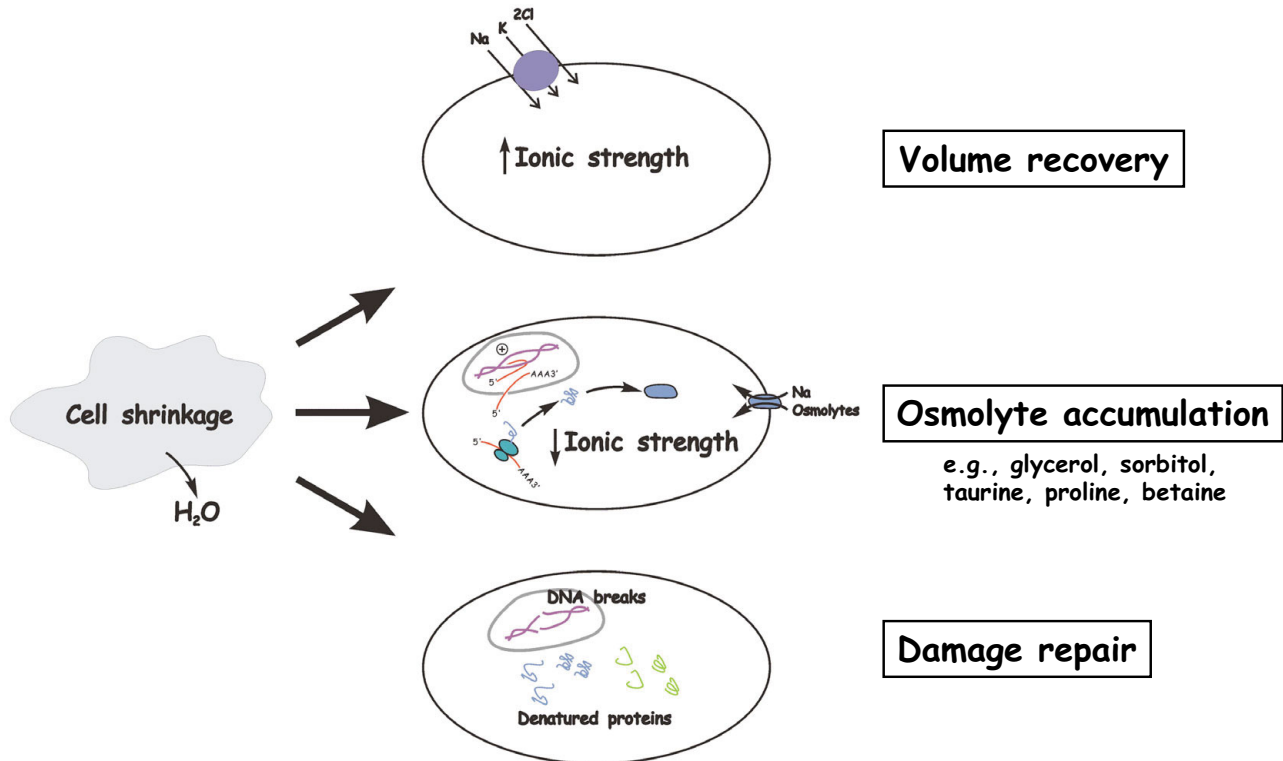


Fig. 1. Cartoon illustrating the hypertonic stress response of animal cells. Exposure to hypertonic media causes rapid water loss and cell shrinkage. Cells respond to shrinkage by activating regulatory volume increase (RVI) salt uptake mechanisms. Osmotically obliged water follows salt uptake and cell volume returns to its original value. Over a period of several hours, cells replace inorganic ions accumulated during RVI with organic osmolytes. Accumulation of organic osmolytes is mediated either by energy-dependent transport from the external medium or by changes in the rates of osmolyte synthesis and degradation. Hypertonic stress typically increases the expression of both organic osmolyte transporters and key enzymes involved in their synthesis. Cells also repair molecular damage including DNA breaks and protein denaturation induced by the initial cell shrinkage and elevation cell inorganic ion levels.

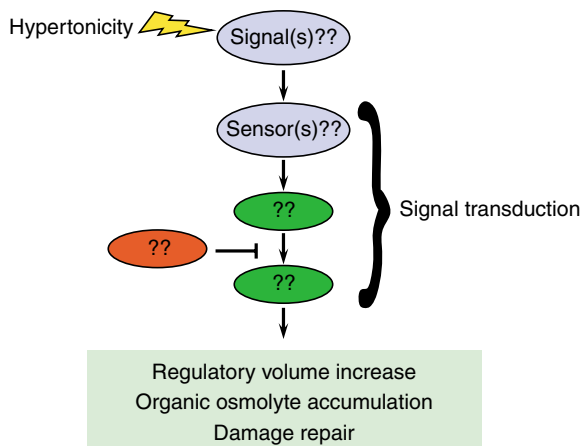


Fig. 2. Cartoon illustrating the steps involved in activation of the cellular hypertonic stress response. In animal cells, the signals by which osmotic stress is detected and the signaling pathways, including inhibitory inputs, that regulate activation of regulatory volume increase (RVI) mechanisms, organic osmolyte accumulation and damage repair are poorly understood. Genome-wide RNAi screening in *C. elegans* suggests that disruption of new protein synthesis and cotranslational protein folding is one signal that activates organic osmolyte accumulation (see Fig. 4) (see also Lamitina et al., 2006).

strong transcriptional upregulation while *gpdh-2* shows a weak, transient increase in expression levels.

To determine whether *gpdh-1* and *gpdh-2* are osmoregulatory effectors, we generated a *gpdh-1;gpdh-2* double knockout worm. The double knockout had no obvious phenotype under normal growth conditions. However, when exposed to hypertonic stress, the worms had greatly reduced glycerol levels, were completely sterile and showed greatly slowed larval development (Lamitina et al., 2006). Transgenic worms expressing *gpdh-1* or *gpdh-2* GFP reporters demonstrated that *gpdh-2* is constitutively expressed in the intestine, hypodermis and excretory cell. *gpdh-1* expression was not detected under normal growth conditions. However, expression was induced in the intestine and hypodermis during hypertonic stress (Lamitina et al., 2006).

The striking on-off behavior of the *gpdh-1* GFP reporter provided an ideal assay for forward and reverse genetic screens. To begin identifying signals and signaling mechanisms that regulate osmoprotective gene expression, we performed a genome-wide RNAi feeding screen using a commercially available RNAi feeding library that contained individual *E. coli* clones engineered to produce double stranded RNA homologous to ~16 000 *C. elegans* genes. Worms were fed

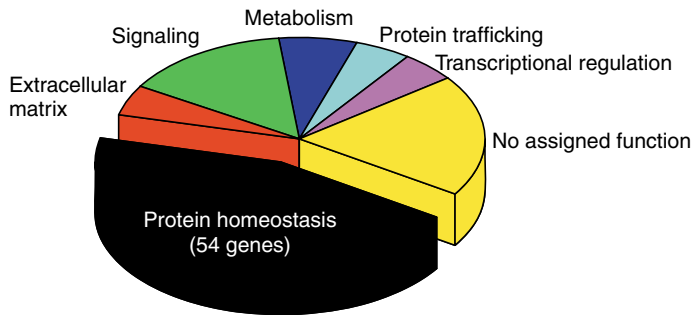


Fig. 3. Pie chart showing distribution of the functional categories of the 122 *rgpd* genes identified by genome-wide RNAi screening in *C. elegans* (Lamitina et al., 2006).

single *E. coli* clones for 3 days and then visually scored for constitutive activation of the *gpdh-1* GFP reporter. This initial screen identified 106 genes whose knockdown induced *gpdh-1* expression in the absence of hypertonic stress. These genes are termed regulators of glycerol-3-phosphate-dehydrogenase (*rgpd*) expression (Lamitina et al., 2006).

Genome-wide RNAi screening results in significant numbers of false negatives (Simmer et al., 2003). To identify additional *rgpd* genes, we queried the *C. elegans* Interactome, a genome-wide protein–protein interaction map composed of 3228 genes and 5685 yeast two-hybrid interactions (Li et al., 2004). Forty-eight of the 106 *rgpd* genes identified in the initial screen were present in the Interactome. These 48 *rgpd* genes interacted with 148 other genes. Of these 148 interacting genes, 124 were represented in our RNAi library. We re-screened these 124 interacting genes by RNAi feeding and identified an additional 16 genes that activated the *gpdh-1* GFP reporter. The Interactome screen thus increased the *rgpd* genes by 15% to 122. This increase is consistent with false negative rates of 10–30% that have been estimated for *C. elegans* genome-wide RNAi screens (Simmer et al., 2003).

rgpd gene functions fell into six defined cellular processes as well as a group of genes with unassigned functions (Fig. 3). Interestingly, the majority (44% or 54/122) of *rgpd* genes fell into a category defined as protein homeostasis. These genes encode proteins required for RNA processing, protein synthesis, protein folding and protein degradation. Protein homeostasis genes function to maintain levels of properly folded and functioning cellular proteins. Inhibition of these genes is expected to increase the levels of damaged cellular proteins. Recent studies (Nollen et al., 2004) support this idea. Wild-type GFP expressed in *C. elegans* muscle cells is distributed uniformly in the cytoplasm. However, modified GFPs containing repeats of glutamine undergo age-dependent aggregation

(Morley et al., 2002). Genome-wide RNAi screening identified 187 genes that function to slow ageing-induced protein aggregation (Nollen et al., 2004). We found that 34 of the 122 *rgpd* genes overlapped with this 187-gene dataset. This is a 24-fold greater overlap than expected by chance alone ($P < 0.001$). Strikingly, 25 of the 34 overlapping genes are predicted to function in RNA processing, protein synthesis, protein folding and protein degradation. Thus, genes that function to prevent protein aggregation also function to inhibit *gpdh-1* expression. When the function of these genes is disrupted, damaged and denatured proteins accumulate in cells and *gpdh-1* expression is increased, leading to glycerol accumulation.

Our results are consistent with a model in which increased levels of damaged or denatured proteins act as signal that triggers osmoprotective gene expression and organic osmolyte accumulation (Fig. 4). Accumulation of organic osmolytes is expected to stabilize protein structure and decrease protein misfolding (e.g. Auton and Bolen, 2005; Ignatova and Gierasch, 2006), which in turn would serve to autoregulate pathway activity.

Interestingly, our experimental observations suggest that *gpdh-1* expression is specifically activated by osmotically induced disruption of new protein synthesis and cotranslational folding rather than by denaturation of existing proteins (Lamitina et al., 2006) (Fig. 4). Such a mechanism would allow cells to discriminate between osmotically induced protein damage and other forms of stress-induced damage. Our

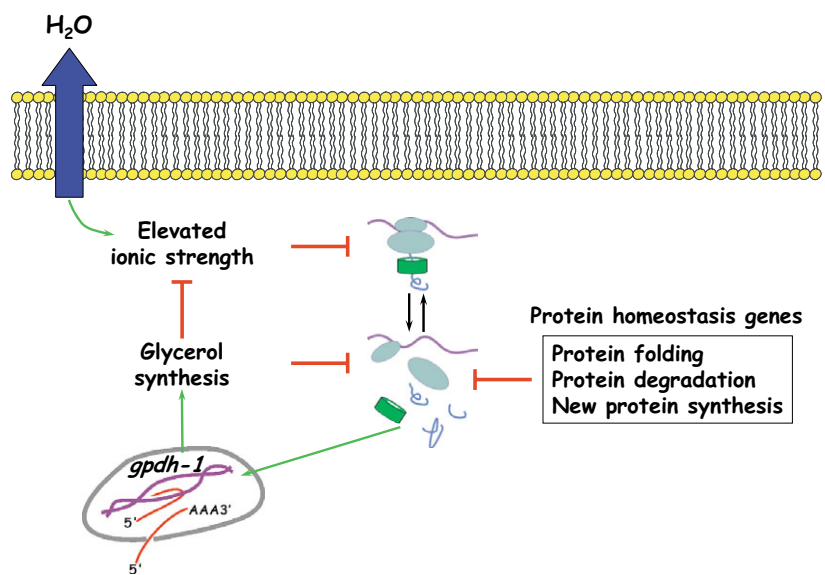


Fig. 4. Model for regulation of osmosensitive gene expression by disruption of protein homeostasis. Hypertonic stress induced water loss causes elevated cytoplasmic ionic strength, which in turn disrupts new protein synthesis and cotranslational protein folding (see Lamitina et al., 2006). Damaged proteins function as a signal that activates *gpdh-1* expression and glycerol synthesis. Glycerol replaces inorganic ions in the cytoplasm and functions as a chemical chaperone that aids in the refolding of misfolded proteins. Loss of function of protein homeostasis genes also causes accumulation of damaged proteins and activation of *gpdh-1* expression.

proposed model is analogous to the unfolded protein response, which is an intracellular signaling and transcriptional/translational program activated by the accumulation of unfolded proteins in the ER lumen that functions to restore ER protein homeostasis (Schroder and Kaufman, 2005).

The *rgpd* genes identified in our RNAi screen represent inhibitory inputs into the signaling pathways that regulate osmoprotective gene expression (see Fig. 2). Loss-of-function mutant worm strains exist for several of these *rgpd* genes and these worms exhibit constitutive *gpdh-1* expression and glycerol accumulation (Lamitina et al., 2006). Importantly, it is now possible to use these mutants and carry out additional RNAi and mutagenesis screens to identify genes that suppress *gpdh-1* expression. Such suppressor genes will almost certainly include components of the signaling pathway that function normally to activate the expression of osmoprotective genes such as *gpdh-1*. Thus, by exploiting the experimental attributes of *C. elegans*, it should eventually be possible to develop an integrated molecular understanding of how an animal cell detects osmotic stress and activates protective mechanisms. It is likely that such mechanisms show strong evolutionary conservation (e.g. Strange et al., 2006). New insights gained from *C. elegans* will therefore undoubtedly provide insights into how more complex organisms including mammals cope with osmotic perturbations.

Conclusions and future perspective

Genome sequences are often referred to as 'molecular' or 'genetic blueprints'. Blueprints are detailed, integrated sets of plans or programs of action that describe how to accomplish a particular task. Given our current level of understanding, a genome is little more than lines of code (i.e. genes) that specify how to synthesize RNA and proteins. Genome sequence must be deciphered into a set of instructions that allow us to understand how organisms are built, how they run and how they interact with their environments. It is the job of the integrative physiologist to decipher the lines of an organism's genetic code into a working blueprint. Genetically tractable organisms such as *C. elegans* provide numerous experimental advantages for defining gene function and integrative biology. Ultimately, however, it will be the integration of insights obtained from many different organisms that provides the deepest understanding of a biological process. In this regard, comparative physiology is very much alive and relevant in the post-genome era.

This work was supported by NIH grants DK51610, DK61168 and GM74229.

References

- Albertson, D. G. and Thomson, J. N. (1976). The pharynx of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **275**, 299-325.
- Ashrafi, K., Chang, F. Y., Watts, J. L., Fraser, A. G., Kamath, R. S., Ahringer, J. and Ruvkun, G. (2003). Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268-272.
- Auton, M. and Bolen, D. W. (2005). Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. USA* **102**, 15065-15068.
- Bargmann, C. I. and Avery, L. (1995). Laser killing of cells in *Caenorhabditis elegans*. *Methods Cell Biol.* **48**, 225-250.
- Berezikov, E., Bargmann, C. I. and Plasterk, R. H. (2004). Homologous gene targeting in *Caenorhabditis elegans* by biolistic transformation. *Nucleic Acids Res.* **32**, e40.
- Bloom, F. E. (2001). What does it all mean to you? *J. Neurosci.* **21**, 8304-8305.
- Boulton, S. J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D. E. and Vidal, M. (2002). Combined functional genomic maps of the *C. elegans* DNA damage response. *Science* **295**, 127-131.
- Brockie, P. J., Mellem, J. E., Hills, T., Madsen, D. M. and Maricq, A. V. (2001). The *C. elegans* glutamate receptor subunit NMR-1 is required for slow NMDA-activated currents that regulate reversal frequency during locomotion. *Neuron* **31**, 617-630.
- Broverman, S., MacMorris, M. and Blumenthal, T. (1993). Alteration of *Caenorhabditis elegans* gene expression by targeted transformation. *Proc. Natl. Acad. Sci. USA* **90**, 4359-4363.
- Christensen, M. and Strange, K. (2001). Developmental regulation of a novel outwardly rectifying mechanosensitive anion channel in *Caenorhabditis elegans*. *J. Biol. Chem.* **276**, 45024-45030.
- Christensen, M., Estevez, A. Y., Yin, X. M., Fox, R., Morrison, R., McDonnell, M., Gleason, C., Miller, D. M. and Strange, K. (2002). A primary culture system for functional analysis of *C. elegans* neurons and muscle cells. *Neuron* **33**, 503-514.
- Cinar, H., Keles, S. and Jin, Y. (2005). Expression profiling of GABAergic motor neurons in *Caenorhabditis elegans*. *Curr. Biol.* **15**, 340-346.
- Colosimo, M. E., Brown, A., Mukhopadhyay, S., Gabel, C., Lanjuin, A. E., Samuel, A. D. and Sengupta, P. (2004). Identification of thermosensory and olfactory neuron-specific genes via expression profiling of single neuron types. *Curr. Biol.* **14**, 2245-2251.
- Davis, M. W., Fleischhauer, R., Dent, J. A., Joho, R. H. and Avery, L. (1999). A mutation in the *C. elegans* EXP-2 potassium channel that alters feeding behavior. *Science* **286**, 2501-2504.
- Estevez, A. Y., Roberts, R. K. and Strange, K. (2003). Identification of store-independent and store-operated Ca²⁺ conductances in *Caenorhabditis elegans* intestinal epithelial cells. *J. Gen. Physiol.* **122**, 207-223.
- Fire, A. (1986). Integrative transformation of *Caenorhabditis elegans*. *EMBO J.* **5**, 2673-2680.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.
- Fox, R. M., Von Steina, S. E., Barlow, S. J., Shaffer, C., Olszewski, K. L., Moore, J. H., Dupuy, D., Vidal, M. and Miller, D. M. (2005). A gene expression fingerprint of *C. elegans* embryonic motor neurons. *BMC Genomics* **6**, 42.
- Franks, C. J., Pemberton, D., Vinogradova, I., Cook, A., Walker, R. J. and Holden-Dye, L. (2002). Ionic basis of the resting membrane potential and action potential in the pharyngeal muscle of *Caenorhabditis elegans*. *J. Neurophysiol.* **87**, 954-961.
- Goodman, M. B., Hall, D. H., Avery, L. and Lockery, S. R. (1998). Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* **20**, 763-772.
- Gunsalus, K. C., Ge, H., Schetter, A. J., Goldberg, D. S., Han, J. D., Hao, T., Berriz, G. F., Bertin, N., Huang, J., Chuang, L. S. et al. (2005). Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861-865.
- Hall, D. H., Lints, R. and Altun, Z. (2006). Nematode neurons: anatomy and anatomical methods in *Caenorhabditis elegans*. *Int. Rev. Neurobiol.* **69**, 1-35.
- Harbinder, S., Tavernarakis, N., Herndon, L. A., Kinnell, M., Xu, S. Q., Fire, A. and Driscoll, M. (1997). Genetically targeted cell disruption in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **94**, 13128-13133.
- Ideker, T., Galitski, T. and Hood, L. (2001). A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343-372.
- Ignatova, Z. and Gierasch, L. M. (2006). Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc. Natl. Acad. Sci. USA* **103**, 13357-13361.
- Jansen, G., Hazendonk, E., Thijssen, K. L. and Plasterk, R. H. (1997). Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat. Genet.* **17**, 119-121.
- Jeon, U. S., Kim, J. A., Sheen, M. R. and Kwon, H. M. (2006). How tonicity

- regulates genes: story of TonEBP transcriptional activator. *Acta Physiol.* **187**, 241-247.
- Jorgensen, E. M. and Mango, S. E.** (2002). The art and design of genetic screens: *Caenorhabditis elegans*. *Nat. Rev. Genet.* **3**, 356-369.
- Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. and Ahringer, J.** (2000). Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**, 2.1-2.10.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. et al.** (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-237.
- Kitano, H.** (2002a). Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr. Genet.* **41**, 1-10.
- Kitano, H.** (2002b). Systems biology: a brief overview. *Science* **295**, 1662-1664.
- Krebs, H. A.** (1975). The August Krogh Principle: 'For many problems there is an animal on which it can be most conveniently studied'. *J. Exp. Zool.* **194**, 221-226.
- Krogh, A.** (1929). Progress in physiology. *Am. J. Physiol.* **90**, 243-251.
- Lamitina, S. T. and Strange, K.** (2004). Transcriptional targets of the DAF-16 insulin signaling pathway protect *C. elegans* from extreme hypertonic stress. *Am. J. Physiol.* **288**, 467-474.
- Lamitina, S. T., Morrison, R., Moeckel, G. W. and Strange, K.** (2004). Adaptation of the nematode *Caenorhabditis elegans* to extreme osmotic stress. *Am. J. Physiol.* **286**, C785-C791.
- Lamitina, T., Huang, C. G. and Strange, K.** (2006). Genome-wide RNAi screening identifies protein damage as a regulator of osmoprotective gene expression. *Proc. Natl. Acad. Sci. USA* **103**, 12173-12178.
- Lang, F., Busch, G. L., Ritter, M., Volkl, H., Waldegger, S., Gulbins, E. and Haussinger, D.** (1998). Functional significance of cell volume regulatory mechanisms. *Physiol. Rev.* **78**, 247-306.
- Lee, S. S., Lee, R. Y., Fraser, A. G., Kamath, R. S., Ahringer, J. and Ruvkun, G.** (2003). A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nat. Genet.* **33**, 40-48.
- Leung, B., Hermann, G. J. and Priess, J. R.** (1999). Organogenesis of the *Caenorhabditis elegans* intestine. *Dev. Biol.* **216**, 114-134.
- Lewis, J. A. and Fleming, J. T.** (1995). Basic culture methods. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (ed. H. F. Epstein and D. C. Shakes), pp. 4-29. New York: Academic Press.
- Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T. et al.** (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* **5657**, 540-543.
- Lockery, S. R. and Goodman, M. B.** (1998). Tight-seal whole-cell patch clamping of *Caenorhabditis elegans* neurons. *Meth. Enzymol.* **293**, 201-217.
- Maricq, A. V., Peckol, E., Driscoll, M. and Bargmann, C. I.** (1995). Mechanosensory signalling in *C. elegans* mediated by the GLR-1 glutamate receptor. *Nature* **378**, 78-81.
- Mello, C. and Fire, A.** (1995). DNA transformation. *Methods Cell Biol.* **48**, 451-482.
- Mello, C. C., Kramer, J. M., Stinchcomb, D. and Ambros, V.** (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.
- Morley, J. F., Brignull, H. R., Weyers, J. J. and Morimoto, R. I.** (2002). The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **99**, 10417-10422.
- Murphy, C. T., McCarroll, S. A., Bargmann, C. I., Fraser, A., Kamath, R. S., Ahringer, J., Li, H. and Kenyon, C.** (2003). Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277-283.
- Nehrke, K.** (2003). A reduction in intestinal cell pH_i due to loss of the *Caenorhabditis elegans* Na⁺/H⁺ exchanger NHX-2 increases life span. *J. Biol. Chem.* **278**, 44657-44666.
- Nelson, F. K. and Riddle, D. L.** (1984). Functional study of the *Caenorhabditis elegans* secretory-excretory system using laser microsurgery. *J. Exp. Zool.* **231**, 45-56.
- Nelson, F. K., Albert, P. S. and Riddle, D. L.** (1983). Fine structure of the *Caenorhabditis elegans* secretory-excretory system. *J. Ultrastruct. Res.* **82**, 156-171.
- Nollen, E. A., Garcia, S. M., van Haften, G., Kim, S., Chavez, A., Morimoto, R. I. and Plasterk, R. H.** (2004). Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation. *Proc. Natl. Acad. Sci. USA* **101**, 6403-6408.
- Pennisi, E.** (2003). Systems biology. Tracing life's circuitry. *Science* **302**, 1646-1649.
- Praitis, V., Casey, E., Collar, D. and Austin, J.** (2001). Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**, 1217-1226.
- Raizen, D. M. and Avery, L.** (1994). Electrical activity and behavior in the pharynx of *Caenorhabditis elegans*. *Neuron* **12**, 483-495.
- Richmond, J. E. and Jorgensen, E. M.** (1999). One GABA and two acetylcholine receptors function at the *C. elegans* neuromuscular junction. *Nat. Neurosci.* **2**, 791-797.
- Richmond, J. E., Davis, W. S. and Jorgensen, E. M.** (1999). UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nat. Neurosci.* **2**, 959-964.
- Riddle, D. L. and Albert, P. S.** (1997). Genetic and environmental regulation of dauer larva development. In *C. elegans II* (ed. D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess), pp. 739-768. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Rual, J. F., Ceron, J., Koreth, J., Hao, T., Nicot, A. S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S. H., Hill, D. E., van den Heuvel, S. et al.** (2004). Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, 2162-2168.
- Rutledge, E., Bianchi, L., Christensen, M., Boehmer, C., Morrison, R., Broslat, A., Beld, A. M., George, A., Greenstein, D. and Strange, K.** (2001). CLH-3, a CIC-2 anion channel ortholog activated during meiotic maturation in *C. elegans* oocytes. *Curr. Biol.* **11**, 161-170.
- Schafer, W. R.** (2005). Deciphering the neural and molecular mechanisms of *C. elegans* behavior. *Curr. Biol.* **15**, R723-R729.
- Schroder, M. and Kaufman, R. J.** (2005). The mammalian unfolded protein response. *Annu. Rev. Biochem.* **74**, 739-789.
- Schwarz, E. M., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Canaran, P., Chan, J., Chen, N., Chen, W. J., Davis, P. et al.** (2006). WormBase: better software, richer content. *Nucleic Acids Res.* **34**, D475-D478.
- Segal, E. and Kim, S. K.** (2003). The modular era of functional genomics. *Genome Biol.* **4**, 317.
- Sen, G. L. and Blau, H. M.** (2006). A brief history of RNAi: the silence of the genes. *FASEB J.* **20**, 1293-1299.
- Sheen, M. R., Kim, S. W., Jung, J. Y., Ahn, J. Y., Rhee, J. G., Kwon, H. M. and Woo, S. K.** (2006). Mre11-Rad50-Nbs1 (MRN) complex is activated by hypertonicity. *Am. J. Physiol.* **291**, F1014-F1020.
- Shen, C., Nettleton, D., Jiang, M., Kim, S. K. and Powell-Coffman, J. A.** (2005). Roles of the HIF-1 hypoxia-inducible factor during hypoxia response in *Caenorhabditis elegans*. *J. Biol. Chem.* **280**, 20580-20588.
- Simmer, F., Moorman, C., Van Der Linden, A. M., Kuijk, E., Van Den Berghe, P. V., Kamath, R., Fraser, A. G., Ahringer, J. and Plasterk, R. H.** (2003). Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* **1**, E12.
- Stinchcomb, D. T., Shaw, J. E., Carr, S. H. and Hirsh, D.** (1985). Extrachromosomal DNA transformation of *Caenorhabditis elegans*. *Mol. Cell Biol.* **5**, 3484-3496.
- Strange, K., Denton, J. and Nehrke, K.** (2006). Ste20-type kinases: evolutionarily conserved regulators of ion transport and cell volume. *Physiology* **21**, 61-68.
- Sulston, J. E. and Horvitz, H. R.** (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110-156.
- Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N.** (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119.
- Tabara, H., Grishok, A. and Mello, C. C.** (1998). RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430-431.
- Timmons, L. and Fire, A.** (1998). Specific interference by ingested dsRNA. *Nature* **395**, 854.
- Timmons, L., Court, D. L. and Fire, A.** (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103-112.
- Viswanathan, M., Kim, S. K., Berdichevsky, A. and Guarente, L.** (2005). A role for SIR-2.1 regulation of ER stress response genes in determining *C. elegans* life span. *Dev. Cell* **9**, 605-615.
- Walhout, A. J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K. C., Schetter, A. J. et al.** (2002).

- Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**, 1952-1958.
- White, J. G., Southgate, E., Thomson, J. N. and Brenner, S.** (1986). The structure of the nervous system of the nematode *C. elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci* **314**, 1-340.
- Williams, D. C., Boulin, T., Ruaud, A. F., Jorgensen, E. M. and Bessereau, J. L.** (2005). Characterization of *Mos1*-mediated mutagenesis in *Caenorhabditis elegans*: a method for the rapid identification of mutated genes. *Genetics* **169**, 1779-1785.
- Yanai, I.** (2003). The integrated world of functional genomics. *Genome Biol.* **4**, 301.
- Yancey, P. H.** (2005). Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* **208**, 2819-2830.
- Yuan, A., Santi, C. M., Wei, A., Wang, Z. W., Pollak, K., Nonet, M., Kaczmarek, L., Crowder, C. M. and Salkoff, L.** (2003). The sodium-activated potassium channel is encoded by a member of the Slo gene family. *Neuron* **37**, 765-773.
- Zhong, W. and Sternberg, P. W.** (2006). Genome-wide prediction of *C. elegans* genetic interactions. *Science* **311**, 1481-1484.
- Zipperlen, P., Fraser, A. G., Kamath, R. S., Martinez-Campos, M. and Ahringer, J.** (2001). Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* **20**, 3984-3992.

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at <http://www.wikipedia.org/>

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology .
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	ch romatin i mmunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray .
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from ch romatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as <i>cis</i> -acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of 'C followed by G' dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C-G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (<i>trans-acting</i> factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called 'gene deserts'.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of <u>gene</u> and <u>chromosome</u> , the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the <i>Hox</i> genes in animals and <i>MADS-box</i> genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	<u>in</u> sertion and <u>de</u> letion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon .
KEGG	The <u>K</u> yoto <u>E</u> ncyclopedia of <u>G</u> enes and <u>G</u> enomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. <i>Drosophila</i> , <i>Caenorhabditis elegans</i> and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3' ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5' and followed by a 3' untranslated region (5' UTR and 3' UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA , siRNA and piRNA . Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA .

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA <i>in vitro</i> . The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3' tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a 'reflectron' by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	<u>RNA-induced silencing complex</u> . A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC .
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR .
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters . Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
<i>trans</i> -acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called 'mobile genetic elements' or 'jumping genes'.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3' or 5' flanking ends of the molecule (i.e. 3' UTR and 5' UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.