Gibbs free energy as a measure of complexity correlates with time within *C. elegans* embryonic development

Samuel H. McGuire¹ · Edward A. Rietman¹ ·
Hava Siegelmann¹ · Jack A. Tuszynski^{2,3}

Received: 22 April 2017 / Accepted: 23 August 2017 / Published online: 19 September 2017
© Springer Science+Business Media B.V. 2017

Abstract We investigate free energy behavior in the nematode *Caenorhabditis elegans* during embryonic development. Our approach utilizes publicly available gene expression data, which gives us a picture of developmental changes in protein concentration and, resultantly, chemical potential and free energy. Our results indicate a clear global relationship between Gibbs free energy and time spent in development and provide thermodynamic indicators of the large-scale biological events of cell division and differentiation.

Keywords Development · Gibbs free energy · Entropy · Protein-protein interaction networks

1 Introduction

Understanding the energetic dynamics during organismal development is a key ingredient to a comprehensive picture of birth defects, developmental disorders, evolution, and the

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10867-017-9469-0>) contains supplementary material, which is available to authorized users.

✉ Jack A. Tuszynski
jackt@ualberta.ca

Edward A. Rietman
erietman@gmail.com

¹ College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA

² Department of Oncology, Cross Cancer Institute, University of Alberta Edmonton, Edmonton, Alberta, Canada

³ Department of Physics, University Of Alberta, Edmonton, Alberta T6G 2J1, Canada

relationship between thermodynamics and life. We could use such a picture, for example, to develop new methodologies for targeted treatments and screenings based on the thermodynamic role of a given protein or enzyme. Already, such methodologies are being used in clinical studies in oncology [1] where correlation has been found between the Gibbs free energy of a particular cancer and the corresponding 5-year survival probability. This paper seeks to take a preliminary step toward extending this methodology in order to gain understanding how developmental thermodynamic behavior may quantify a systems level by examining the nematode *Caenorhabditis elegans*.

We create a ‘time lapse’ of protein concentration and chemical potential energy on an organismal level by imposing time-course RNA-seq data over an organism-wide protein pathway known as an *interactome*. The expression level of a given gene then serves as a proxy for protein concentration and, as a result, chemical potential energy. This approach balances a relatively high-resolution image of the organism with accessibility in terms of available data. One would reasonably expect that free energy in this context, then, does not encompass all energetic components of the organism’s cellular environment but, rather, provides an approximate lower bound. With this in consideration, we note that high-throughput protein concentration time-course data would, naturally, provide a more direct analysis and that a statistical treatment of enzymes and non-coding RNA would allow us to further specify the thermodynamic behavior of the organism.

This concept we propose is first to connect Gibbs free energy and development on a systems level, though Rietman et al. in [1] displayed significant relationships between free energy and various characteristics of cancerous tumors. Specifically, correlations have been demonstrated to exist between Gibbs free energy and 5-year cancer survival probability ($R = -0.72$) as well as Gibbs free energy and cancer stage ($R = -0.99$). The latter aspect is strongly suggestive that there should be a relationship between Gibbs free energy and development stage in general.

The organization of this paper is as follows: the following section contains background information and justifications for the use of *C. elegans*, the Gibbs free energy of a network and how it can be calculated using an interactome and RNA-seq data. After the background material, we present the main results, which indicate a positive correlation between the magnitude of Gibbs free energy and embryonic development.

2 Background

2.1 *Caenorhabditis elegans* primer

C. elegans is a self-fertilizing, compost-dwelling nematode which often serves as a model organism in genetics, neuroscience, development and other areas of biology [2]. This is due in part to its practicality as a research tool, a characteristic generated by a combination of factors such as the ability to self-reproduce, a transparent epidermis and a relatively low cost to maintain and grow [2]. This practicality has allowed scientists to determine at a high resolution many of its biological mechanisms, such as cell lineage and neural architecture and genome [3]. In addition to having a complete picture of the *C. elegans* genome, its interactome is well-studied, which allows for an increased accuracy in this context. For our purposes, *C. elegans* has a dual interest in that its complexity may be sufficient to extend our results to larger metazoa and also that it provides an abundance of high-quality, publicly available data. We restrict our attention to hermaphroditic samples, which serves as the overwhelming majority of *C. elegans* offspring.

The life-cycle of a *C. elegans* hermaphrodite consists of five main phases: an embryonic period, which after hatching leads to four larval stages denoted in increasing maturity L1, L2, L3 and L4 and an adult phase in which reproduction is possible [2]. Within embryonic development, there are two stages: *in utero*, lasting approximately 150 min. The mother then lays the embryo outside, beginning the *ex utero* phase, lasting approximately 9 ± 3.5 h, where variations are generated by temperature and other environmental factors. Within the embryo, worms develop their general body plan as well as the capacity for movement and pharyngeal pumping [2]. These physiological developments occur in two stages (whose timelines are independent of the *in* and *ex utero* phases just mentioned). The first stage is *proliferation*, during which cell division occurs and lasts until roughly 350 min after fertilization. The second phase, *morphogenesis*, marks a temporary halt in cell division and cell energy (metabolism) is redirected towards differentiation. It is during *morphogenesis* that the worm develops organs and its usual body plan. Hatching marks the end of the *ex utero* phase [4], which, as mentioned, occurs between 700 and 900 min after fertilization. The beginning of larval development requires the presence of food. During larval development, a molt is characteristic of transitional periods. The approximate length of each larval stage is as follows: L1 $\rightarrow$ 12 h, L2 $\rightarrow$ 8 h, L3 $\rightarrow$ 8 h, L4 $\rightarrow$ 10 h [2]. Organisms develop their nervous and reproductive systems during the larval phases. After the L4 stage, roughly 50 h after conception, the worm enters the reproductively-capable adult stage, generally laying eggs 10 h later [5].

2.2 Protein-protein interaction networks

We choose to employ an organism-wide protein-protein interaction network—also referred to as an *interactome*—as the interactional structure to model the *C. elegans* cell environment. The methodologies for creating and characteristics of interactome networks are reviewed in [6]. Formally, an interactome is an undirected graph G consisting of an edge set E and a vertex set V . Within the edge set E there are no multi-edges. A vertex or node $v \in V$ represents a protein, and each edge $\langle i, j \rangle \in E$ represents a possible interaction between the proteins i and j . Similarly, nodes adjacent to a given node i , denoted as $Adj(i)$, comprise the set of i 's possible interactions. The cardinality of this set, $|Adj(i)|$, is referred to as the *degree* of i . The adjacency matrix of an undirected graph is matrix $A \in \{0, 1\}^{|V| \times |V|}$ such that $A_{i,j} = 1$ iff $\langle i, j \rangle \in E$.

The details and source for the interactome used in the study can be found in the Data Sources section. Within this interactome $|V| = 6176$ and $|E| = 178151$. As a disclaimer, the nature of the interactome can only provide us an approximate picture of the proteomic environment of *C. elegans* because of the incompleteness inherent to this stage of development in high-throughput proteomic analysis and technology. This fact requires us to be careful in our analysis of results and contributes to the notion that these results are both approximate and indicative of lower bounds, as mentioned in the introduction. We now examine the relationship between gene expression and protein concentration and the prospect of using the former to gain insight into the dynamics of the latter.

2.3 mRNA and protein concentration during *C. elegans* development

The gradients of protein concentrations within a cell can provide us with a basic notion of chemical potential, serving as the foundation of free energy. High throughput protein concentration methods are not always available and are certainly more expensive than transcriptome data, so we use transcriptome data as a proxy for protein concentration. This allows a simple mapping into the interactome network. Moreover, embryonic protein

analysis via mass spectrometry, which will be necessary for an investigation of development, is limited due to the discrepancy between embryo protein mass and the protein mass requirements for MS [7]. This limitation is surmountable but presents a bottleneck for available data. To empirically justify the choice to use mRNA abundance and thus gene expression as a proxy for protein abundance, both [8] and [9] demonstrate a Pearson correlation coefficient ranging from 0.36 to 0.84 between RNA expression and protein concentration, measured via mass spectrometry, across various species. Other research has also reported a favorable correlation between transcription signal and proteomic information [10, 11]. On the other hand, the correlations between mRNA and protein abundance have had mixed results within varying contexts in other studies using similar techniques [7, 12–14]. In Peshkin et al. [7], a conceptual explanation for the poor correlation may be due to the limiting role mRNA concentration plays in the rate protein synthesis. Given a small mRNA/protein ratio, the small mRNA level will cause the changes in translation rate to minimally affect the significantly higher protein level. Further, Smits et al. [14] suggest that variations in protein abundance are due to post-transcriptional modifications and degradation. As an example—due to [7]—a low correlation between mRNA and protein abundance in embryonic development may be due to excessive protein synthesis and successive protein stockpiling in order to make efficient use of maternal resources. Maintaining these ‘stock-piled’ proteins can be done via post-transcriptional modification in order to prevent their premature degradation. This consideration further contributes to the role of our results as lower bounds—mRNA abundance is a limiting factor for protein level and can provide a simplified picture of a protein gradient in the cellular environment. More detailed, dynamical data on protein concentration—ideally without the use of conceptual proxy—will allow an even higher-resolution examination of energy dynamics during development.

For definitional clarity, a set of expression data is formed by doing an RNA-seq expression analysis on a set of related samples. In our case, different samples correspond to different parts of the developmental timeline. In the same vein, an expression set can be considered as a matrix $B \in \mathbb{R}^{n \times s}$ where n is the number of gene expressions being queried (which, in the case of *C. elegans*, is $\sim 20,000$) and s is the number of samples. If we impose a time-ordering on the samples, then a row vector $\vec{v} \in B$ details the changes in expression as time progresses in the course of the experiment. To create a weighted network, we simply assign each node its corresponding expression vector, thus defining a ‘family’ of interaction networks for which each member corresponds to a specific time-point within the experiment.

It is important to note that there are not one-to-one correspondences between the genome annotations in our interactome and those in our data sets. This is because they employ different versions of the WormBase genome [15–17]. To remediate this, we take the intersection between the interactome gene set and the expression data gene set and concern ourselves with the subgraph induced by such intersection. Let S_I and S_E be the gene sets of the interactome and expression data, respectively, and $S = S_I \cap S_E$. In addition, let G_S be the subgraph of G induced by S . Then, in Section 4, we will use the notion of *completeness* of both the protein (node) and interaction (edge) sets in reference to the percentages determined by $\frac{|V_{G_S}|}{|V_G|}$ and $\frac{|E_{G_S}|}{|E_G|}$, respectively.

2.4 Gibbs free energy in interaction networks

Our methods are the same as those used in [1]. As we have discussed in the section prior, the interactome’s structure represents observed protein-protein interactions. Importantly, though, a given protein is not participating in all possible interactions at any single given

point in time. It is a reasonable description of the organism's protein environment that copies of the given protein are present in different locations both within a particular cell and between the cells within an organism. This description is equivalent to the notion of an *ensemble* from statistical mechanics, where each copy of the given protein and the interactions such copy participates in represents a possible state of the protein environment. With this, we can assume the given protein, its copies and the interactions between them form an ideal gas mixture.

With this assumption, we can use the common expression for the Gibbs free energy within a chemical system:

$$G_{total} = \sum_i c_i \mu_i \quad (1)$$

where c_i is the concentration of the i th member of the chemical system and μ_i is its chemical potential. A given gene within a $\log_2$ -normalized expression set will have an expression value that falls in the interval $[0, 20]$, though this can vary depending on the experiment. To recover the notion of molar concentration from these expression values, we normalize it into the interval $[0, 1]$. To do so, we restrict our attention to the expression values of a given sample—this can be seen as a column vector within the expression set matrix—and let e_{max} and e_{min} be the maximum and minimum expression value across genes profiled within the sample. Then, given the expression of the i th protein e_i , we define the concentration of the i th protein c_i :

$$c_i = \frac{e_i - e_{min}}{e_{max} - e_{min}} \quad (2)$$

which is clearly bounded by 0 and 1 when e_i is equal to e_{min} and e_{max} , respectively. This strategy allows highly down-regulated genes to take the concentration value 0 and highly up-regulated genes to take the concentration value 1, as opposed to, for example, an e_i/e_{max} normalization.

The notion of chemical potential is restricted within a network because the potential energy of a given member is restricted by the quantity of interactions that can occur within the set of interactions it is able to participate in. The assumption of an ideal gas mixture further allows us to assign a simple chemical potential to a given member of the network:

$$\mu_i = \ln \left[\frac{c_i}{\sum_{j \in Adj(i)} c_j} \right] \quad (3)$$

which is a restricted version of the entropy of mixing. By substitution, we get the expression for the Gibbs free energy within a network:

$$G_{total} = \sum_i c_i \ln \left[\frac{c_i}{\sum_{j \in Adj(i) \cup \{i\}} c_j} \right]. \quad (4)$$

In general, the free energy *magnitude* of the system increases as the reactant concentration increases. Notice that for a given protein, $\frac{c_i}{\sum_{j \in Adj(i) \cup \{i\}} c_j} \leq 1$ with equality occurring when the i th protein is 'functionally' disconnected, meaning it is connected only to nodes with concentration 0. Furthermore, the logarithm is bounded from above by 0 and approaches $-\infty$ with an increasing neighborhood concentration. It follows, then, that G_{total} will never be positive. As is the case with Gibbs free energy in its usual chemical context, a negative value indicates an exergonic system in which input energy is not required for interactions to occur and therefore captures the likelihood that spontaneous reactions occur. Accordingly, when we refer to an increase in free energy *magnitude*, it indicates, by definition, an increase in thermodynamic favorability and spontaneity.

3 Data sources

3.1 Expression sets

The expression set used was accessed through PubMed with accession number GSE60755 [17]. We took raw read count data and normalized it by taking the logarithm with one additional pseudo-read. That is, we let $e = \log(r+1)$ where e is the new expression value, r is the raw read count and the logarithm is base 2. GSE60755 examines embryonic development up to hatching.

3.2 Interactome

The interactome used was acquired from Dr. Marc Vidal's Center for Cancer Systems Biology website and was constructed using a combination of methods, some of which are detailed by Simonis et al. in [18]. In particular, the interactome is a union of high-throughput yeast two-hybrid networks, literature-curated networks, orthologous interaction networks and correlational techniques with phenotypic and microarray analysis. It is available for download via [15].

3.3 Software

Network constructions and analyses were completed using the Python package `networkx` and Cytoscape.

4 Results & discussion

We created an interactome whose nodes are weighted by normalized expression values using the process outlined above with the materials mentioned. The relative completenesses of this augmented interactome is very high, indicating that the majority of the original interactome's structure is preserved. In particular, GSE60755 has node- and edge-completeness of 0.94 and 0.89, respectively. The completeness, as defined above, with respect to the edge and node sets is a measure of how much information is common across both the given expression data set and the interactome. Such a measure is necessary because it is not guaranteed that all information should be preserved due to different releases of the *C. elegans* genome used in gathering either the interactome or expression data. With this weighted interactome, we calculated the Gibbs free energy using the concept from (1)–(4). We then averaged the Gibbs free energy across replicate samples at a common time point to create an 'average' *C. elegans* with respect to the expression experiment, creating, then, a single free energy vector for each expression set. Figure 1 shows a time plot of Gibbs free energy resulting from the averaging process.

These results indicate that the Gibbs free energy absolute value clearly *decreases* as the *C. elegans* develops within the embryo. In physics, this decrease can be interpreted as reflecting a decreased *stability* of the system, or, equivalently, resilience to change affected by external work. However, it should be also noted one of the properties of living systems is entropy reduction due to their metabolic activity. In these calculations we have not included data on the organism's metabolic rates but it can be safely assumed that they increase as the system develops and matures. Therefore, this is consistent with the trend toward entropy reduction (Gibbs free energy increase) correlating with an organism's

Fig. 1 Gibbs free energy vs. time for embryonic development. This is computed using the methods outlined above using the expression set with accession number GSE60755 (available at [17]) and a yeast two-hybrid *C. elegans* interactome, available at [15]. The results indicate a decrease in the Gibbs free energy magnitude across this timespan, which physically corresponds to a decrease in stability and potential work within the system, as well as an increase in disorder

robustness and greater differentiation that requires energy inputs. Incidentally, this is an opposite trend to that observed in cancer progression, which implies de-differentiation and entropy increase, hence Gibbs free energy decrease. These relationships can be directly seen as entropy and Gibbs free energy which can be compared using the following fundamental thermodynamic relation:

$$G = H - TS \quad (5)$$

where G is Gibbs free energy, H is the enthalpy, T is the temperature and S is the entropy of the thermodynamic system. At constant temperature, our results indicate that entropy is decreasing over the course of embryonic development. Most importantly, complexity is increasing as a result of continued differentiation. Modern literature considers life to be anomalous from a thermodynamic perspective due to the tendency for life-forms to exist in a thermodynamically stable state (without large fluctuations in entropy) while far from equilibrium (low entropy). In this context, free energy is what performs stabilizing work. In order to fully understand our results, post-hatching thermodynamic analysis of metabolism needs to be completed.

The biological shift in the embryogenic process from the first stage, proliferation, to the second stage, morphogenesis, might also explain the behavior. Morphogenesis marks the stage during which cells begin to differentiate and cell division halts. Thus, within a network setting, proteins responsible for differentiation become specialized and thus no longer represent degree-based ‘hubs’, decreasing the effect that concentration of any given protein has on the total Gibbs free energy. Figure 2 illustrates the behavior of the Gibbs free energy during these two phases of embryonic development. To further justify this claim on the two

distinct patterns within our results for GSE60755, we split up the results into the time periods during which proliferation and morphogenesis take place. According to [2], this shift takes place during roughly 350 min into development. This line of inquiry tells us that proliferation has a linear fit with a modest correlation, $R = 0.2195$, but a relatively inactive trend: the Gibbs free energy over this time span varies between -6500 and -4500 without a clear global pattern. Morphogenesis, on the other hand, shows a slightly improved

Fig. 2 [Upper] Gibbs free energy during the morphogenesis phase of *C. elegans* embryonic development. This phase corresponds to high cell division and no cell differentiation activity [Lower] Gibbs free energy during the proliferation phase of *C. elegans* embryonic development. This phase corresponds to low cell division and high cell differentiation activity

correlation, $R = 0.4286$, with the additional feature that the slope of the fit during morphogenesis much more clearly indicates a global behavior of decreasing free energy magnitude, decrease in entropy and increase in complexity: in fact, the slope is *3 times greater* than the slope of the fit during proliferation. This indicates the possibility that cell division, which halts during morphogenesis, acts as a thermodynamically stabilizing process.

4.1 Degree preserving randomization

To examine the significance of the interactome's architecture in our results, we subjected the interactome to degree preserving randomizations. Figure 3 depicts the Gibbs free energy vs. time averaged over common time points for 100 randomized graphs. The close similarity of these results with the original interactome indicates the possible importance of the role of a node's degree. As further examination, we generated 100 Erdős-Rényi random graphs with gene labels and put them through the same analysis, which is also depicted in Fig. 3. The similarity across all three network models demonstrates a peculiar resilience of Gibbs free energy to the topology of its underlying network. It should be noted, though, that the strength of the linear correlation in the case of a fully randomized Erdős-Rényi graph is weaker and there is a larger variance in the results. Thus this overall behavior might be due to statistical events within the expression data itself, such as overall abundance of reactants corresponding to high expression levels.

These global results suggest that we should locate the topological features of the interactome, which would allow us to distinguish between the three models: a real biological network, a network with identical topology but no biological meaning, and a completely randomized network. We can begin to address this question in local detail by examining the *homology networks* of the interactome and its randomized counterparts. The homology network of a given graph is a subgraph whose nodes consist of the t highest-energy genes where t is a given parameter known as the *threshold*. The homology network of a graph with threshold 48, then, would be the subgraph consisting of the 48 most energetic nodes. Such an object corresponds to a filtration of the energy landscape within a cell generated by the concentration of proteins and their interactional structure. In this setting, each time-point in development has a corresponding homology network for a fixed threshold t , since each time-point represents an energy landscape distinct from the last.

We can use these homology networks to directly distinguish the random Erdős-Rényi graphs from the interactome. In particular, the original interactome has homology networks that are densely connected, possessing, on average, about half of the possible $\binom{n}{2}$ edges. The Erdős-Rényi graphs have Gibbs homology networks which are generally quite disconnected, which indicate multiple 'peaks' of the associated energy landscapes. Example homology networks from the original interactome and the randomized networks are pictured in Fig. 4.

On the other hand, the topologies of the Gibbs homology networks do not directly distinguish the interactome from its degree-preserved counterparts. In particular, the homology networks, in all cases, are connected graphs with a high level of completeness in terms of average node degree vs. graph size. This tells us that both the original and degree-preserving interactomes possess energy landscapes with concentrated peaks which favor high-degree nodes. This similarity can be justified as follows:

Gibbs free energy (as defined above) provides a measure of concentration versus neighborhood concentration. Degree preserving randomizations maintain the *size* of the neighborhood, but do not necessarily preserve the neighborhood concentration, as the specific proteins and their associated concentrations are randomized. With this in mind, it

Fig. 3 [Upper] Gibbs free energy vs. time for embryonic development in degree-preserving randomized graphs generated from the *C. elegans* interactome available via [15]. We generated 100 random graphs using an endpoint-shuffle algorithm and the results depicted are mean Gibbs free energy values across those 100 graphs. [Lower] Gibbs free energy vs. time for embryonic development in Erdős-Rényi random graphs generated with the same node- and edge-set size as the *C. elegans* interactome used in our initial results. The results depict the mean Gibbs free energy across 100 graphs. Both figures show a striking similarity with our original results, which indicates that network architecture may be a less significant feature in our model than the statistical events (corresponding to changes in protein concentration) that occur within the gene expression data across development

Fig. 4 Threshold-48 Gibbs homology networks of GSE60755 and DPR45 at time-point 109 (out of 139; 520 min after fertilization and arbitrarily chosen for illustrative purposes). [Upper] Original interactome homology network with, $|E| = 1106$ and $|V| = 48$. [Middle] DPR interactome homology network. $|E| = 793$, $|V| = 48$. [Lower] ER homology network, where the 23 connected components consisting of a single node are excluded. $|E| = 25$, $|V| = 48$

might seem curious that the empirically observed interactome does not appear to provide a thermodynamic difference in terms of the global Gibbs free energy. But the nature of a degree-preserving randomization means that high-degree nodes within a graph with power-law distribution will often remain connected, whereas low-degree nodes will often be connected to entirely different nodes. See SI Fig. 1 to verify the power-law distribution of our interactome and thus its degree-preserved counterparts. Perhaps it is the degree-distribution that creates a smooth energy landscape which is more biologically feasible, as evidenced by the difference in the average connected components between the interactome (and its degree-preserved analogues) and the Erdős-Rényi homology networks. Indeed, the degree distribution and the concentration (and, by definition, Gibbs free energy) are what ultimately inform the homology networks' topologies, meaning that the homology networks don't encode information that can distinguish an interactome from a degree-preserved counterpart in this setting. This would further indicate that statistical events in the expression data have a larger bearing on the global character of the Gibbs free energy, but the degree-distribution is what encodes protein functionality and biological feasibility.

5 Conclusions

We have demonstrated a correlation between the Gibbs free energy and time in *C. elegans* development. Of particular note is the negative slope of this correlation (when considering the magnitude of the Gibbs free energy), which can be partially explained by differentiating between the distinct embryogenic processes of proliferation and morphogenesis. Doing so indicates that cell division plays a fundamental role in Gibbs free energy changes over the course of development and, correspondingly, the stability of a thermodynamic system.

This hypothesis can play a satisfying philosophical role as well—the Gibbs free energy is associated with the ability to maintain order in the face of increasing entropy. If we think of organisms that have particularly sophisticated and effective mechanisms by which they maintain order, we can naturally associate it with complexity. This paper has indicated the possibility that Gibbs free energy is linked to the interplay between cell division and differentiation, which allows us to tie the concept of organismal complexity to the temporal exchange of those mechanisms. Drawing connections between a larger range of metabolic events may allow us to distinguish between the thermodynamic roles of anabolic and catabolic processes within an organism's metabolism. We have noted that the developmental trend in terms of Gibbs free energy increase over time is opposite to that seen in tumor progression, which is consistent with the biologically opposite processes of differentiation in embryonal development and de-differentiation in tumorigenesis. All life forms use metabolic energy to reduce entropy whose natural tendency in non-living matter is to increase with time [19].

Tumor growth occurs with a tendency to increase entropy compared to normal tissue, which endows it with a competitive advantage vis a vis survival under harsh environmental conditions (hypoxia, acidity, etc.). For a complete analysis of thermodynamic features of both cancer cells and embryonal development, additional data in the form of metabolic rates and their changes are needed, which will be addressed in a future publication.

Further improvements can be made to both the model and the data as higher resolution data becomes available. In terms of the model, we can use temporal networks that more accurately reflect the cellular environment at any given point in time and thus give us a more accurate picture of the network thermodynamics. Of course, more accurate interactomes will also provide us with more accurate results. Furthermore, data that allow us to

probe the energetic events within the cell in which protein interactions do not participate would also give us a more detailed understanding of thermodynamics. If we treat Gibbs free energy as a distribution across time, it is an interesting possibility to measure the differences between evolutionarily distinct organisms by using traditional statistical methods to measure the distance between two probability distributions.

Acknowledgements We acknowledge (NSF)/ECCS-1533693 NSC-FO: Col “Individual Variability in Human Brain Connectivity, Modeling Using Multi-scale Dynamics Under Energy Constraints”. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation. This research was also supported by the Office of Naval Research and we acknowledge the award number N00014-15-1-2126.ONR

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest

References

1. Rietman, E.A., Platig, J., Tuszynski, J.A., Lakka Klement, G.: Thermodynamic measures of cancer: Gibbs free energy and entropy of protein–protein interactions. *J. Biol. Phys.* **42**(3), 339–350 (2016). <https://doi.org/10.1007/s10867-016-9410-y>
2. Altun, Z.F., Hall, D.H.: Introduction. In: *WormAtlas*. <https://doi.org/10.3908/wormatlas.1.1> (2009)
3. Corsi, A.K., Wightman, B., Chalfie, M.A.: Transparent window into biology: a primer on *Caenorhabditis elegans*. *WormBook* http://www.wormbook.org/chapters/www_celegansintro/celegansintro.html (2015). Accessed 3 March 2017
4. Bucher, E.A., Seydoux, G.C.: Gastrulation in the nematode *Caenorhabditis elegans*. *Semin. Dev. Biol.* **5**, 121–130 (1994). <https://doi.org/10.1006/sedb.1994.1016>
5. Byerly, L., Russell, R.L., Cassada, R.C.: The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. *Dev. Biol.* **51**, 23–33 (1976). [https://doi.org/10.1016/0012-1606\(76\)90119-6](https://doi.org/10.1016/0012-1606(76)90119-6)
6. Vidal, M., Cusick, M.E., Barabási, A.-L.: Interactome networks and human disease. *Cell* **144**, 986–998 (2011)
7. Peshkin, L., et al.: On the relationship of protein and mRNA dynamics in vertebrate embryonic development. *Dev. Cell* **35**, 383–394 (2015)
8. Maier, T., Güell, M., Serrano, L.: Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009)
9. Greenbaum, D., Colangelo, C., Williams, K., Gernstein, M.: Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117 (2003)
10. Wilhelm, M., et al.: Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014)
11. Kim, M.-S., et al.: A draft map of the human proteome. *Nature* **509**, 575–581 (2014)
12. Munoz, J., et al.: The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **7**, 1–13 (2011)
13. Grün, D., et al.: Conservation of mRNA and protein expression during development of *C. elegans*. *Cell Rep.* **6**, 565–77 (2014)
14. Smits, A.H., et al.: Global absolute quantification reveals tight regulation of protein expression in single *Xenopus* eggs. *Nucleic Acids Res.* **42**, 9880–9891 (2014)
15. Center for Cancer Systems Biology: Worm Interactome Version 8. http://interactome.dfci.harvard.edu/C_elegans/index.php?page=download. Accessed 3 March 2017
16. Levin, M., Hashimshony, T., Wagner, F., Yanai, I.: Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* **22**, 1101–1108 (2012)
17. Levin, M., et al.: The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016)
18. Simonis, N., et al.: Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat. Methods* **6**, 47–54 (2009)
19. Schrödinger, E.: What is life?: with mind and matter and autobiographical sketches. Cambridge University Press, Cambridge (1992)