Supplementary information: Balance of osmotic pressures determines the nuclear-to-cytoplasmic volume ratio of the cell

Dan Deviri[1] and Samuel A. Safran[1]

[1]Department of Chemical and Biological Physics, Weizmann Institute of Science, Rehovet 76100, Israel

1 Gibbs-Donnan calculation for the osmotic pressure of chromatin counterions

The limit we consider leads to an upper bound for the osmotic pressure of the chromatin counterions. In particular, we assume that the counterions are not localized near the chromatin fiber and their concentration is uniform within the nucleoplasm. Localization and inhomogeneity of the ion concentration lowers their contribution to the osmotic pressure. In addition, we consider the limiting case where there are no cytoplasmic confined charges, and the cytoplasm volume is infinitely larger than the nucleoplasm volume; as we show in the next appendix, this limit, which is appropriate for isolated nuclei, results in an upper bound for the net outward osmotic pressure of chromatin counterions. The concentrations of the different ions are then determined by two conditions: (1) Equality of the electrochemical potentials of the ions in the cytoplasm and nucleus. This potential, in the limit of a dilute solution, is the sum of entropic contribution equal to the log of the concentration of that ion species and a constant electrostatic term whose contribution is of opposite sign for positive and negative charges, $k_B T \log \left(c_i^{\pm} v_0\right) \pm e \nu \Phi_i$; where $k_B T$ is the thermal energy, $c_i^{\pm}$ is the concentration of positive/negative monovalent ions in compartment $i = c, n$ (cytoplasm or nucleoplasm), v_0 is the ion volume, taken to be the same for cations and anions, e is the electron charge, ν is the valence of the ion ($\nu = 1$ for monovalent ions), and Φ_i is the electrostatic potential in compartment i. Equality of the electrochemical potential is an equilibrium condition, which is appropriate for ions that can freely diffuse through the NE which has only passive channels connecting the nucleoplasm and cytoplasm and no ion pumps. (2) The nucleoplasm and the cytoplasm are both electro-neutral. The electrostatic energy cost of mesoscopically charged cytoplasm or nucleus is very large and thus suppresses such distributions [1].

In our minimal model for osmotic pressure of the counterions, the nucleoplasm comprises three types of solutes: (1) positive monovalent ions, (2) negative monovalent ions, and (3) chromatin; the cytoplasm, which lacks confined charges, contains only the ions. We denote the respective concentrations of positive and negative monovalent ions in the cytoplasm by c_c^+ and c_c^- and their respective concentrations in the nucleoplasm by c_n^+ and c_n^-. The net, negative charge concentration of the chromatin (in units of electron charges) in the nucleoplasm is written as c_{ch}^-. From condition (1) of equality of electrochemical potentials for the two types of ions, the concentrations of the monovalent ions are related by $k_B T \log \left(c_c^+/c_n^+\right) = -k_B T \log \left(c_c^-/c_n^-\right)$. This relation is equivalent to the following, simpler relation, $c_c^+ c_c^- = c_n^+ c_n^-$. We note that because the chromatin cannot cross the NE, it does not reach electrochemical equilibrium and it does not satisfy these equations.

Condition (2) for the overall electro-neutrality of the cytoplasm and nucleoplasm, leads to the following two equations, $c_c^+ = c_c^-$ and $c_n^+ = c_n^- + c_{ch}^-$, for each compartment respectively. As mentioned above, we work in the limit where the volume of the cytoplasm is much larger than that of the nucleus, so that ion exchange with the nucleus does not significantly change the ion concentration in the cytoplasm. This simplification allows us to approximate the cytoplasmic concentrations of the positive and negative ions as the intracellular salt concentration $c_c^+ = c_c^- \approx c_{salt}$; corrections to this approximation are not expected to change the order of

magnitude of our pressure estimates. The equations $c_c^+ = c_c^- \approx c_{salt}$, $c_n^+ = c_n^- + c_{ch}^-$, and $c_c^+ c_c^- = c_n^+ c_n^-$ relate the concentrations of the monovalent ions in the nucleoplasm to the salt concentration and the concentration of chromatin counterions:

$$c_n^+ = \frac{\sqrt{\left(c_{ch}^-\right)^2 + 4c_{salt}^2} + c_{ch}^-}{2} \tag{1}$$

$$c_n^- = \frac{\sqrt{\left(c_{ch}^-\right)^2 + 4c_{salt}^2} - c_{ch}^-}{2} \tag{2}$$

In the ideal gas limit, the net contribution of the monovalent ions to the osmotic pressure is simply the difference between the total concentrations of all the monovalent ions in the nucleoplasm and all the monovalent ions in the cytoplasm, multiplied by $k_B T$:

$$p_c = k_B T \left(c_n^+ + c_n^- - 2c_{salt}\right) \approx k_B T \frac{\left(c_{ch}^-\right)^2}{4c_{salt}} \tag{3}$$

where the rightmost expression results from expansion of $\sqrt{\left(c_{ch}^-\right)^2 + 4c_{salt}^2}$ for the biological case that $c_{salt} \gg c_{ch}^-$ (see main text). Substituting the relation between the net chromatin charge concentration, the net chromatin charge, and the nuclear volume, $c_{ch}^- = N_{ch}/V_n$, into Eq. 3 results in Eq. 4 in the main text.

We note that our simple model of the nuclear electrostatics that considers only one type of positive, monovalent ions, one type of negative, monovalent ions, and chromatin, does not capture the true and varied composition of the cellular cytoplasm and nucleoplasm. The cytoplasm and nucleoplasm contain a multitude of ion species with various valencies, as well as charged proteins. However, the majority of the positive and negative ions in the cells are potassium ions and phosphate ions [2], respectively. This justifies our simplification that the positive and negative ions are all of the same type. Furthermore, while our choice of approximating all the ions as monovalent may affect the detailed, numerical value of the estimate, it is not expected to change its order of magnitude. In contrast, the cytoskeleton which is also negatively charged is not included in our estimate either. Its effect on the osmotic pressure can only lower our upper bound estimate. This is because the charge of the cytoskeleton is negative and confined to the cytoplasm, so that its contribution to the osmotic pressure difference between the nucleoplasm and cytoplasm only reduces our estimated contribution of the negatively charged chromatin that is confined to the nucleus. If the concentration of the cytoskeletal charge is comparable to that of the chromatin, it may significantly reduce the osmotic pressure difference. Therefore, the pressure predicted by 3 is an upper bound to the pressure contributed by the chromatin counterions. This is further explicated in the appendix "Detailed model for counterions osmotic pressure" below, which focuses on a detailed, numerically-solvable model for the osmotic pressure of counterions that considers the presence of confined charges in the cytoplasm.

2 Detailed model for counterions osmotic pressure

In this appendix, we estimate the counterion osmotic pressure difference between the nucleoplasm and cytoplasm when we consider the finite cytoplasmic volume and confined charges it contains such as the negatively charged cytoskeletal proteins [3] and membrane phospholipids. Therefore, this appendix complements the derivation presented in Appendix "Gibbs-Donnan calculation for the osmotic pressure of chromatin counterions" above and discussed in the main text for isolated nuclei (corresponding to an infinite cytoplasmic volume). The model presented in this Appendix is more detailed than the previous one and results in less analytically tractable, but nonetheless solvable, set of equations that predict the number of ions in each compartment. The theory is based on the Gibbs-Donnan equilibrium [1] whose equations provide an estimate for the difference of the nuclear and cytoplasmic osmotic pressures of the counterions. In this more detailed model, we quantitatively show that the case of isolated nuclei presented in the previous Appendix lead to an even larger estimate of the net, outward counterion pressure. Since we seek an upper bound to this quantity, the idealization of isolated nuclei, and Eq. 3 that results from this simplification, are appropriate.

As explained in the previous Appendix, since small ions can cross the NE, the electrochemical potentials of each of the two charge species of monovalent ions in the two compartments are equal in chemical equilibrium [1]. This implies that the products of the concentrations of positive and negative monovalent ions in the two compartments are also equal

$$\frac{N_c^+ N_c^-}{V_c^2} = \frac{N_n^+ N_n^-}{V_n^2} \tag{4}$$

where V_c and V_n are the respective volumes of the cytoplasm and nucleoplasm, N_c^+ and N_c^- are the number of positive and negative ions in the cytoplasm, respectively, and similarly, N_n^+ and N_n^- are respectively the number of positive and negative ions in the nucleoplasm. Furthermore, minimization of the electrostatic energy indicates that each of the two compartments is approximately electroneutral, which is described by the following two equations:

$$N_n^- + N_{bn} = N_n^+ \tag{5}$$
$$N_c^- + N_{bc} = N_c^+ \tag{6}$$

where N_{bn} and N_{bc} are the net number charges that are confined to the nucleoplasm (e.g., chromatin) and cytoplasm (e.g., cytoskeleton and lipid bilayers), respectively.

Eqs. 4, 5, and 6 are supplemented by the condition the the number of negative, monovalent ions in both the nucleus and cytoplasm are conserved

$$N_c^- + N_n^- = N^- \tag{7}$$

where N^- is the total number of monovalent anions, which is related to the cellular salt concentration c_{salt}, and total volume of the cell by the following equation

$$N^- = c_{salt}\left(V_n + V_c\right) \tag{8}$$

Importantly, conservation of the total number of cations is automatically satisfied by the conservation of number of anions (Eq. 7) and the electroneutrality of the two compartments (Eqs. 5 and 6), thus does not contribute an additional equation.

Since we are interested in the effect that cytoplasmic, confined charges and the finite volume of the cytoplasm have on the net outward pressure of the counterions, we must consider two additional parameters compared to the case of isolated nuclei. The first is the nucleoplasm-to-cytoplasm volume ratio $\alpha = V_n/V_c$; since the nuclear volume is typically about 8% of the cell volume [4], we consider a range for α between 0.1 and 1. The second parameter we introduce accounts for the number of confined cytoplasmic charges. For simplicity of presentation, we consider that charge as originating from one protein species, which we choose to be actin due to its abundance in the cytoplasm. Therefore, the second parameter is the concentration of actin which typically is in the range of 10-100 µM in non-muscle cells [2, 5]. Since each actin monomer has a negative charge of $Z = 13$ electron charges [3], the concentration of actin c_{actin} is related to the total confined negative charge in the cytoplasm by $N_{bc} = c_{actin} Z V_c$.

To find the net pressure of the counterions for specific values of α and c_{actin}, we numerically solve all the equations above for the positive and negative ion concentrations in the nucleoplasm and cytoplasm, using the relations $\alpha = V_n/V_c$ and $N_{bc} = c_{actin} Z V_c$. To compare the estimates of this more detailed model with the simplified one discussed in the main text, we use the same values for the nuclear volume V_n, nuclear confined charge N_{bn}, and intracellular salt concentration c_{salt} we use in the main text for human cells, $V_n = 1000$ µm³, $N_{bn} = N_{ch} \approx 5.8 \cdot 10^9$, and $c_{salt} = 200$ mM. Then, the numbers of monovalent ions in each compartment, calculated by solving the equations above, are used to estimate the net outward osmotic pressure of the counterions, Δp, in the limit of an ideal solution, by the following formula:

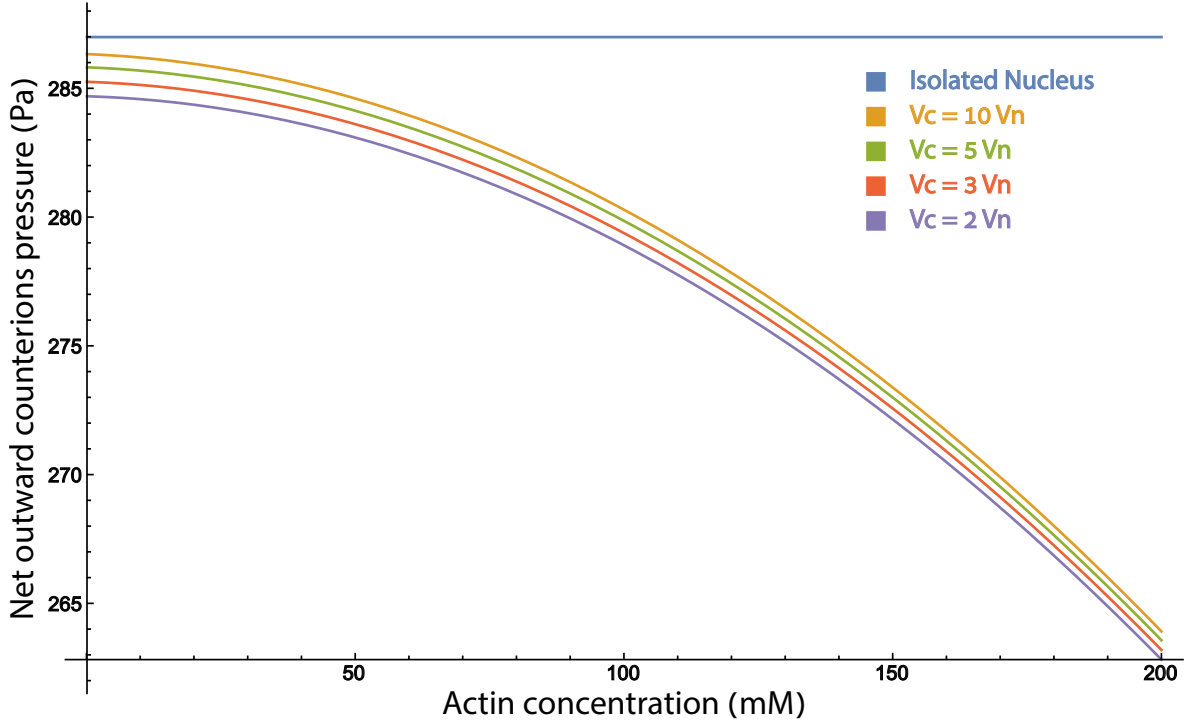$$\Delta p = k_B T \left(\frac{N_n^- + N_n^+}{V_n} - \frac{N_c^- + N_c^+}{V_c}\right) \tag{9}$$

3

Figure 1: Net outward pressure of the counterions as predicted by a more detailed model that takes into account the finite volume of the cytoplasm and the presence of confined charges within it. Here, the volume of the nucleus $V_n \approx 1000 \ \mu m^3$ and confined nuclear charges $N_{cn} \approx 5.8 \cdot 10^9$ are appropriate for human nuclei [2, 6]. As is shown in the figure and indicated in Eq. 9 for Δp, reducing the number of charges confined to the cytoplasm (e.g., by decreasing c_{actin}) increases the net outward pressure, as does increasing the volume of the cytoplasm. Consistent with the arguments presented in the main text, the net outward pressure predicted by Eq. 3 for isolated nuclei is an upper bound for the net outward pressure for the entire range of parameters checked; importantly, a similar plot with parameters appropriate for yeast nuclei, $V_n \approx 17 \ \mu m^3$ and $N_{bn} = 2.5 \cdot 10^7$ shows the same trend (data not shown).

Fig. S1 shows the difference of the nuclear and cytoplasmic pressures, Δp, as a function of the concentration of confined proteins, c_{actin}, for several values of α as well as the estimate for the case of isolated nuclei in Eq. 3 above. Indeed, as argued in the main text, we find that Eq. 3 for the case of an infinite cytoplasm provides an upper bound for the net outward osmotic pressure of the counterions.

3 Localized protein pressure

Studies show that in X. laevis oocytes, most of the cellular proteins are preferentially localized either to the cytoplasm or to the nucleoplasm [7]. Each of those proteins contributes to the net osmotic pressure of the compartment to which it is preferentially localized, a contribution proportional to its concentration in the compartment. This concentration is difficult to measure, since fractionation protocols of nucleoplasm and cytoplasm many times use detergents which perforate the NE, mixing the proteins of the nucleoplasm and cytoplasm [7]. For that reason, concentrations of cellular proteins are commonly measured and calculated in a cell-averaged manner, using total amount of proteins of a specific species in the entire cell and the volume

of the entire cell. The cell-averaged concentration of a localized protein is lower than its concentration in the compartment to which it is preferentially-localized and larger than its concentration in the other compartment. Therefore, the use of cell-averaged protein concentrations may potentially lead to underestimation of the net contribution of preferentially-localized proteins to the osmotic pressure. In this Appendix, we relate the compartment-specific osmotic pressures due to preferentially-localized proteins to the cell-averaged concentrations of those proteins for the important case discussed in the main text: The NE is relaxed and the osmotic pressure due to the localized proteins is the dominant contribution to the pressures of the two compartments.

In the cell, nucleocytoplasmic transport of proteins (and RNA molecules) through the nuclear pore complexes (NPCs) occurs via a shuttling mechanism that is coupled to existing concentration gradients of GDP-Ran and GTP-Ran across the NE [8, 9, 10]. These gradients are maintained by active enzymatic activity that differs in the cytoplasm and nucleoplasm. Due to the active nature of these enzymatic activities, the steady-state of the shuttling mechanism may be out-of-equilibrium, thus violating detailed balance in which the import and export rates are equal (e.g., due to diffusion that would occur for a passive channel). Thus, non-equilibrium activity implies that the import rate, k_i, is different than the export rate, k_e. In steady-state, in any time interval, the amount of any specific protein species that is imported must be equal to its amount that is exported. In the framework of first order kinetics, this condition relates the differing import and export rates to the cytoplasmic, c_c, and nucleoplasmic, c_n, protein concentrations: $k_i c_c = k_e c_n$. This implies that the ratio of the nucleoplasmic and cytoplasmic concentrations of the protein is equal to the ratio of the import and export rates, $c_n/c_c = k_i/k_e$. Since enzymatic transport activity can result in unequal import and export rates, the protein will be enriched in one compartment relative to the other.

Different protein species have different affinities to the molecular components involved in the nucleoplasmic shuttling mechanism, which suggests that each species that is actively transported may have different import and export rates and thus a different relative enrichment in one of the two compartments [9, 10]. We denote the relative enrichment of protein i by κ_i, defined as the ratio of the concentrations of the protein in the nucleoplasm and in the cytoplasm, $\kappa_i \equiv c_{i,n}/c_{i,c}$; if $\kappa_i > 1$ then protein i is actively transported and preferentially-localized to the nucleus, if $\kappa_i < 1$, then protein i is actively transported and preferentially-localized to the cytoplasm. For simplicity, we consider here only proteins that are actively transported to one of the compartments so that $\kappa_i, \kappa_j \neq 1$ for any protein in the summations below.

Using the definition of the relative enrichment, we write the net contribution of the preferentially-localized proteins to the osmotic pressures of the cytoplasm or the nucleoplasm as p_c and p_n, respectively; these are related to the concentrations of the proteins and their relative enrichment by:

$$p_c = k_B T \sum_i (c_{i,c} - c_{i,n}) = k_B T \sum_i (1 - \kappa_i) c_{i,c} \tag{10}$$

$$p_n = k_B T \sum_j (c_{j,n} - c_{j,c}) = k_B T \sum_j (\kappa_j - 1) c_{j,c} \tag{11}$$

where the subscript i denotes summation over different protein species preferentially-localized to the cytoplasm and the subscript j denotes summation over different protein species preferentially-localized to the nucleoplasm. We denote the total, cell-averaged concentrations of all the proteins that are preferentially-localized to the cytoplasm and all the proteins that are preferentially-localized to the nucleoplasm as $\bar{c}_c$ and $\bar{c}_n$, respectively. These total, cell-averaged concentrations are expressed using their compartment-specific concentrations and the free volumes of the compartments as follows:

$$\bar{c}_c = \frac{\sum_i ((V_c - V_{c,m}) c_{i,c} + (V_n - V_{n,m}) c_{i,n})}{(V_c - V_{c,m}) + (V_n - V_{n,m})} = \frac{\sum_i (\alpha c_{i,c} + c_{i,n})}{\alpha + 1} = \frac{\sum_i (\alpha + \kappa_i) c_{i,c}}{\alpha + 1} \tag{12}$$

$$\bar{c}_n = \frac{\sum_j ((V_c - V_{c,m}) c_{j,c} + (V_n - V_{n,m}) c_{j,n})}{(V_c - V_{c,m}) + (V_n - V_{n,m})} = \frac{\sum_j (\alpha c_{j,c} + c_{j,n})}{\alpha + 1} = \frac{\sum_j (\alpha + \kappa_j) c_{j,c}}{\alpha + 1} \tag{13}$$

where as before, subscripts i and j respectively denote summation over protein species that are preferentially-localized to the cytoplasm and nucleoplasm. The free volumes of the cytoplasm and nucleoplasm are respec-

tively denoted as $V_c - V_{c,min}$ and $V_n - V_{n,min}$, and α denotes their ratio, $\alpha \equiv (V_c - V_{c,m}) / (V_n - V_{n,m})$. Using Eqs. 12 and 13, we write the total, cell-averaged concentration of all localized proteins (cytoplasmic and nucleoplasmic) as

$$\bar{c}_c + \bar{c}_n = \frac{\sum_i (\kappa_i - 1) c_{i,c} + \sum_j (\kappa_j - 1) c_{j,c}}{\alpha + 1} + \sum_i c_{i,c} + \sum_j c_{j,c} \tag{14}$$

For the case we consider in this Appendix, that the NE is relaxed and the dominant contributions to the inward and outward pressures originate from localization of proteins, the osmotic pressures of the preferentially-localized proteins in the two compartments, p_c and p_n, are equal due to mechanical equilibrium (see main text). Using Eqs. 10 and 11, and the definition of the relative enrichments, the equality of the two pressures results in the following two relations:

$$\sum_i (\kappa_i - 1) c_{i,c} = -\sum_j (\kappa_j - 1) c_{j,c} \tag{15}$$

$$\sum_i c_{i,n} + \sum_j c_{j,n} = \sum_i c_{i,c} + \sum_j c_{j,c} \tag{16}$$

When Eq. 15 is substituted into Eq. 14 for the total average, cellular concentrations of all localized proteins (cytoplasmic and nucleoplasmic), we find that $\bar{c}_c + \bar{c}_n = \sum_j c_{j,c} + \sum_i c_{i,c}$. This indicates that the total, cell-averaged concentrations of all proteins that are actively transported, is equal to the sum of the cytoplasmic concentrations of the proteins that are preferentially-localized to the cytoplasm, and the cytoplasmic concentrations of the proteins that are preferentially-localized to the nucleoplasm. In turn, Eq. 16 indicates that this is also equal to the sum of the nucleoplasmic concentrations of the proteins that are preferentially-localized to the nucleoplasm and the nucleoplasmic concentrations of the proteins that are preferentially-localized to the cytoplasm. As we now show, this conclusion is also true for the simple case considered in the main text, where there are N_c^l proteins that are <u>completely</u> (rather than preferentially) localized to the cytoplasm and N_n^l that are <u>completely</u> localized to the nucleoplasm. In this case, our conclusion indicates that from the equality of osmotic pressures of these proteins (for the case that the NE is relaxed), $N_c^l / (V_c - V_{c,m}) = N_n^l / (V_n - V_{n,m})$, the following two identities follow, $\left(N_c^l + N_n^l\right) / (V_c - V_{c,m} + V_n - V_{n,m}) = N_n^l / (V_n - V_{n,m})$ and $\left(N_c^l + N_n^l\right) / (V_c - V_{c,m} + V_n - V_{n,m}) = N_c^l / (V_c - V_{c,m})$. These identities are proven by the following simple algebraic steps:

$$\frac{N_c^l}{(V_c - V_{c,m})} = \frac{N_n^l}{(V_n - V_{n,m})} \Rightarrow$$

$$\frac{N_c^l}{N_n^l} = \frac{(V_c - V_{c,m})}{(V_n - V_{n,m})} \Rightarrow$$

$$\frac{N_c^l + N_n^l}{N_n^l} = \frac{(V_c - V_{c,m}) + (V_n - V_{n,m})}{(V_n - V_{n,m})} \Rightarrow$$

$$\frac{N_c^l + N_n^l}{(V_c - V_{c,m}) + (V_n - V_{n,m})} = \frac{N_n^l}{(V_n - V_{n,m})} = \frac{N_c^l}{(V_c - V_{c,m})}$$

This calculation indicates that if about 80% of the proteins are preferentially-localized to either compartment [7], and the total osmotic pressure of all proteins is about 10 kPa [11], then the osmotic pressures in the cytoplasm and nucleoplasm due to localization of proteins is of the order of 8 kPa. As explained in detail in the main text, this 8 kPa estimate is an upper bound for the contribution of the localized proteins to the net inward and outward pressures.

4 Relation to nucleocytoplasmic transport

In this appendix, we generalize the values used in the main text, N_c^l and N_n^l, which represent the number of solutes completely localized to the cytoplasm and nucleoplasm, respectively. The main text focuses on the simple case that there is only one species of macromolecular solute that is found only in the cytoplasm and another species of solute that is found only in the nucleoplasm. Here, we generalize these values to the biologically realistic scenario where there are multiple macromolecular solutes (proteins and RNA molecules) which are preferentially-localized rather than completely-localized. Namely, they are found in both the cytoplasm and nucleoplasm, although in different concentrations, which are determined by the characteristics of the active nucleocytoplasmic transport mechanisms. As explained in the section "Localized protein pressure" above, the nucleoplasmic and cytoplasmic concentrations of each protein i that is actively transported are different. This is quantified by the relative enrichment κ_i, defined as the ratio between the nucleoplasmic and cytoplasmic concentrations of the protein $\kappa_i \equiv c_{i,n}/c_{i,c}$; if $\kappa_i > 1$ then solute i is actively transported and is preferentially localized to the nucleus, if $\kappa_i < 1$, then solute i is actively transported and is preferentially localized to the cytoplasm. Here, we consider only solutes that are actively transported to one of the compartments so that $k_i \neq 1$ for any solute i in the summations below.

In the case that the NE is relaxed, mechanical balance is characterized by Eq. 1 in the main text with $\sigma_n = 0$, which indicates that the osmotic pressure of all the proteins localized on each of the two sides of the NE is equal (for the case that these pressures are the dominant contributions to the net inward and outward pressures). Since the small solutes that are not transported have the same concentration in the nucleoplasm and cytoplasm, the contribution of each small solute species to the net osmotic pressures of either compartment is zero. The osmotic pressure balance of the remaining large proteins that are actively transported is written in the approximation of an ideal solution as:

$$k_B T \sum_i c_{i,n} \quad = \quad k_B T \sum_i c_{i,c} \tag{17}$$

where the summation is over all of the proteins that are actively transported so that their concentrations in the cytoplasm and nucleoplasm are unequal.

In addition to Eq. 17, the number of molecules of each protein species is conserved, so that $c_{i,c}$ and $c_{i,n}$ satisfy the following relation

$$N_i = (V_c - V_{c,m})\,c_{i,c} + (V_n - V_{n,m})\,c_{i,n} \tag{18}$$

where N_i is the total, cellular copy number of protein i, and $V_c - V_{c,m}$ and $V_n - V_{n,m}$ are the respective free volumes of the cytoplasm and nucleoplasm (see main text). We note that over long enough timescales, proteins are produced and degraded in vivo, thus our use of the conservation of proteins is only appropriate when the experimental timescales are much shorter than the protein production and degradation timescales (which are of the order of 10 hours [2, 12]).

Eq. 18, together with the definition of the relative enrichment, $\kappa_i \equiv c_{i,n}/c_{i,c}$, can be used to express the concentrations of the proteins in the cytoplasm and nucleoplasm using the free volumes of the two compartments and the total, cellular copy numbers of the different protein species:

$$c_{i,c} \quad = \quad \frac{N_i}{(V_c - V_{c,m}) + \kappa_i\,(V_n - V_{n,m})} \tag{19}$$

$$c_{i,n} \quad = \quad \frac{\kappa_i N_i}{(V_c - V_{c,m}) + \kappa_i\,(V_n - V_{n,m})} \tag{20}$$

Substituting Eqs. 19 and 20 into the condition for the mechanical equilibrium of the relaxed NE, Eq. 17 and multiplying by $V_n - V_{n,m}$, results in an implicit equation for the ratio of the free volumes $(V_c - V_{c,m})\,/\,(V_n - V_{n,m})$

$$\sum_i \frac{\kappa_i N_i}{\frac{(V_c - V_{c,m})}{(V_n - V_{n,m})} + \kappa_i} \quad = \quad \sum_i \frac{N_i}{\frac{(V_c - V_{c,m})}{(V_n - V_{n,m})} + \kappa_i} \tag{21}$$

7

Solving Eq. 21 to find the ratio between the free volumes $(V_c - V_{c,m}) / (V_n - V_{n,m})$ as an explicit function of the different copy numbers, N_i and the relative enrichment coefficients, κ_i, is not possible. However, this equation implies that this ratio is a constant that is related to the N_i-s and the κ_i-s. Furthermore, Eq. 21 allows us to derive an important identity that relates the ratio of the free volumes to the total number of proteins in each compartment (those that are preferentially localized to the compartment and other species in that compartment that are preferentially localized to the other compartment), N_c^l and N_n^l; notably, those correspond to the quantities used in the minimal model presented in the main text. To do this, we first express N_c^l and N_n^l as the following functions of the free volumes, protein copy numbers N_i, and relative enrichment coefficients κ_i:

$$N_c^l = \sum_i (V_c - V_{c,m}) c_{i,c} = \sum_i \frac{(V_c - V_{c,m}) N_i}{(V_c - V_{c,m}) + \kappa_i (V_n - V_{n,m})} \qquad (22)$$

$$N_n^l = \sum_i (V_n - V_{n,m}) c_{i,n} = \sum_i \frac{\kappa_i (V_n - V_{n,m}) N_i}{(V_c - V_{c,m}) + \kappa_i (V_n - V_{n,m})} \qquad (23)$$

Multiplication of Eq. 21 by $(V_c - V_{c,m}) / (V_n - V_{n,m})$, followed by substitution of Eqs. 22 and 23 into the resulting product, results in the relation

$$\frac{V_c - V_{c,m}}{V_n - V_{n,m}} = \frac{N_c^l}{N_n^l} \qquad (24)$$

This is the same conclusion that we arrived at in the main text, where we used N_c^l and N_n^l to denote a simple case where only one species is completely localized in each compartment. Therefore, the detailed calculations presented in this appendix justify the use of N_c^l and N_n^l in the minimal model presented in the main text. The rest of the results presented in the main text all rely on Eq. 24 which we derived here rigorously, and are thus equally justified.

5 Cytoplasmic and nuclear volumes in the non-ideal limit

In the main text, we solved Eqs. 9 and 10 that represent mechanical equilibrium across the NE and the plasma membrane, for the ideal solution limit where the minimal volumes of the cytoplasm and nucleoplasm are negligible compared to their respective volumes and for relaxed NE and plasma membrane ($\sigma_n = 0$ and $\sigma_p = 0$). This yielded expressions 11 and 12 in the main text that relate the volumes of the cytoplasm and nucleoplasm to the extra-cellular solute concentration and nubmers of completely-localized and non-localized solutes. These equations indicated that the ratio between the volume of the nucleoplasm and the cytoplasm (NC ratio) depends only on the ratio of the numbers of completely-localized proteins in the nucleoplasm and in the cytoplasm. However, the corrections to the ideal limit which we used to determine the volumes of the cytoplasm and nucleoplasm may be significant, since the minimal volume of the cell, and presumably the minimal volumes of the separate compartments, may reach tens of percent of their total volumes [6]. In this Appendix, we now include the effects of the minimal volumes (steric, excluded volume of the various solutes). We show that the prediction regarding the NC ratio for $\sigma_n = 0$ and $\sigma_p = 0$ is also valid in the non-ideal limit, as long as the volumes of the non-diffusive complexes in the compartments are small compared with the total volumes of the compartments.

In the non-ideal limit, the free volumes of the two compartments, $V_c - V_{c,m}$ and $V_n - V_{n,m}$, replace the total volumes of the compartments in Eqs. 11 and 12 in the main text, which results in the following equations

$$V_c - V_{c,m} = \frac{N_c^l}{C} \left(1 + \frac{N}{N^l} \right) \qquad (25)$$

$$V_n - V_{n,m} = \frac{N_n^l}{C} \left(1 + \frac{N}{N^l} \right) \qquad (26)$$

where N is the total number of non-localized, small solutes in the cell, N_c^l and N_n^l are the number of solutes that are completely-localized to the cytoplasm and nucleoplasm, respectively, C is the concentration of solutes in the extra-cellular environment, and $N^l = N_c^l + N_n^l$ is the total number of localized solutes in the entire cell.

In experiments, the total volumes of the two compartments, V_n and V_c, are measured rather than their free volumes, $V_n - V_{n,m}$ and $V_c - V_{c,m}$. This motivates us to express the minimal volumes of the two compartments using the numbers of different solutes defined above, in order to rewrite Eqs. 25 and 26 as explicit functions of the observables, V_c and V_n. The minimal volumes of the two compartments are the total volume of all their non-solvent components, including small, non-localized solutes, large, localized solutes, and non-diffusive structures within each the two compartments (e.g. cytoskeleton and chromatin). To account in an algebraically simple manner for the different sizes of localized vs. non-localized solutes [13] in our minimal model, we consider different average volumes for these two types of molecules: The volumes of the various small, non-localized solutes (including ions and small molecules) are taken to be equal, v_s, while the volumes of the various large, localized solutes are also taken to be equal, v_ℓ, where $v_\ell > v_s$. The average volumes of the different solutes allow us to write the minimal volumes as $V_{c,m} \approx v_s N_c + v_\ell N_c^l + V_{c,s}$ and $V_{n,m} \approx v_s N_n + v_\ell N_n^l + V_{n,s}$, where respectively: N_c and N_n are the numbers of non-localized solutes in the cytoplasm and nucleoplasm, N_c^l and N_n^l are the total numbers of localized solutes in the cytoplasm and nucleoplasm, and $V_{c,s}$ and $V_{n,s}$ are the volumes of the non-diffusive structures in the cytoplasm and nucleoplasm. Substituting the expressions for the minimal volumes and the ratio of the free volumes (Eq. 10 in the main text) into main-text Eqs. 6 and 7 for N_c and N_n, and the resulting expressions into Eqs. 25 and 26, we derive the following expressions for the cytoplasmic and nuclear volumes, V_c and V_n

$$V_c = N_c^l \left(1 + \frac{N}{N^l} \right) \left(\frac{1}{C} + \bar{v} \right) + V_{c,s} \tag{27}$$

$$V_n = N_n^l \left(1 + \frac{N}{N^l} \right) \left(\frac{1}{C} + \bar{v} \right) + V_{n,s} \tag{28}$$

where N is the total number of non-localized solutes in the cell, $N^l = N_c^l + N_n^l$ is the total number of localized solutes in the cell, and $\bar{v} = \left(v_s N + v_\ell N^l \right) / \left(N + N^l \right)$ is the average volume of a solute in the cell. We observe that if $V_{c,s} \ll V_c$ and $V_{n,s} \ll V_n$, the ratio of the nuclear and cytoplasmic volumes is well-approximated by the ideal solution limit (see main text) and is again equal to the ratio, N_n^l / N_c^l, of the numbers of their respectively localized proteins. This is the typical biological situation in which the total volumes of the non-diffusing structures in the cytoplasm and nucleoplasm, $V_{c,s}$ and $V_{n,s}$, are much smaller than the total volumes of their respective compartments. The volume of the cytoskeleton, which is presumably the largest non-diffusing structure in the cytoplasm, is only few percent of the volume of the cytoplasm [14]. Similarly, the calculations below of the total volume fraction of (bare) chromatin in human and S. pombe nuclei, based on the structural properties of DNA and histone proteins, show that it is of the order of one percent of the volume of the nucleoplasm.

The volume of chromatin in the nucleus is the sum of the total volumes of the DNA and its associated proteins. Since the most abundant and largest DNA-binding, protein complexes are the histones, the total volume of the chromatin is well-approximated by the sum of the DNA and histone volumes. From crystallographic structural measurements of DNA and histone octamers, the physical dimensions of the two biomolecules are known: DNA occupies a volume approximated by a cylinder of radius of 1 nanometer whose contour length is 0.34 nm multiplied by the number of base pairs [15]. This results in a DNA volume which is about 1 nm^3 per base pair. A histone octamer occupies a volume approximated by a cylinder whose diameter is 6.5 nm and height 6 nm [15], so that the volume of each histone octamer is ~200 nm^3. The average density of histones in a genome is typically one histone octamer per ~200 base pairs [15], which indicates that the presence of histones contributes about 1 nm^3 per base pair of DNA, similar to the contribution of the DNA itself. Therefore, the total volume of the chromatin (in nm^3) of a eukaryotic organism is about twice the size of its genome in base pairs.

In human cells, the diploid genome is $6.4 \cdot 10^9$ base pairs long [2], which implies that the total volume of the chromatin is $\sim 12.8 \cdot 10^9$ nm^3. In S. pombe cells, the diploid genome is $27.6 \cdot 10^6$ base pairs long [16],

which similarly implies that the total volume of the chromatin is $\sim 55.2 \cdot 10^6$ nm^3. In order to calculate the volume fraction of chromatin in the nuclei of the two organisms, we divide the chromatin volumes of human and S. pombe cells by the respective volumes of their typical nuclei, $\sim 1000 \mu$m^3 [6] and $\sim 17 \mu$m^3 [17]. We find that the volume fractions of chromatin in the two species is $\sim 1.3\%$ for humans and $\sim 0.32\%$ for S. pombe, both are of the order of one percent or less. Of course, this does not include the water of hydration, screening cloud of counterions or other proteins that might bind to chromatin.

In conclusion, the low volume fractions of the cytoskeleton and chromatin, presumably the largest non-diffusive cellular structures, allow us to approximate the NC ratio calculated from Eqs. 27 and 28 in the non-ideal solution limit as the ratio N_n^l/N_c^l, similar to the one calculated in the main text for the ideal limit.

References

[1] J Th Overbeek. The donnan equilibrium. *Prog. Biophys. Biophys. Chem*, 6(1):57–84, 1956.

[2] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015.

[3] Paul A Janmey, David R Slochower, Yu-Hsiu Wang, Qi Wen, and Andrejs Cēbers. Polyelectrolyte properties of filamentous biopolymers and their consequences in biological fluids. *Soft matter*, 10(10):1439–1449, 2014.

[4] Helena Cantwell and Paul Nurse. Unravelling nuclear size control. *Current genetics*, 65(6):1281–1285, 2019.

[5] Toren Finkel, Julie A Theriot, Kirk R Dise, Gordon F Tomaselli, and Pascal J Goldschmidt-Clermont. Dynamic actin structures stabilized by profilin. *Proceedings of the National Academy of Sciences*, 91(4):1510–1514, 1994.

[6] Ming Guo, Adrian F Pegoraro, Angelo Mao, Enhua H Zhou, Praveen R Arany, Yulong Han, Dylan T Burnette, Mikkel H Jensen, Karen E Kasza, Jeffrey R Moore, et al. Cell volume change through water efflux impacts cell stiffness and stem cell fate. *Proceedings of the National Academy of Sciences*, 114(41):E8618–E8627, 2017.

[7] Martin Wühr, Thomas Güttler, Leonid Peshkin, Graeme C McAlister, Matthew Sonnett, Keisuke Ishihara, Aaron C Groen, Marc Presler, Brian K Erickson, Timothy J Mitchison, et al. The nuclear proteome of a vertebrate. *Current biology*, 25(20):2663–2671, 2015.

[8] Ronen Benjamine Kopito and Michael Elbaum. Reversibility in nucleocytoplasmic transport. *Proceedings of the National Academy of Sciences*, 104(31):12743–12748, 2007.

[9] Sanghyun Kim and Michael Elbaum. Enzymatically driven transport: a kinetic theory for nuclear export. *Biophysical journal*, 105(9):1997–2005, 2013.

[10] Sanghyun Kim and Michael Elbaum. A simple kinetic model with explicit predictions for nuclear transport. *Biophysical journal*, 105(3):565–569, 2013.

[11] Ron Milo. What is the total number of protein molecules per cell volume? a call to rethink some published values. *Bioessays*, 35(12):1050–1055, 2013.

[12] Roy D Dar, Brandon S Razooky, Abhyudai Singh, Thomas V Trimeloni, James M McCollum, Chris D Cox, Michael L Simpson, and Leor S Weinberger. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, 109(43):17454–17459, 2012.

[13] Einat Grossman, Ohad Medalia, and Monika Zwerger. Functional architecture of the nuclear pore complex. *Annual review of biophysics*, 41:557–584, 2012.

[14] Qi Wen and Paul A Janmey. Polymer physics of the cytoskeleton. *Current Opinion in Solid State and Materials Science*, 15(5):177–182, 2011.

[15] Helmut Schiessel. The physics of chromatin. *Journal of Physics: Condensed Matter*, 15(19):R699, 2003.

[16] V Wood, R Gwilliam, M-A Rajandream, M Lyne, R Lyne, A Stewart, J Sgouros, N Peat, J Hayles, S Baker, et al. The genome sequence of schizosaccharomyces pombe. *Nature*, 415(6874):871–880, 2002.

[17] Frank R Neumann and Paul Nurse. Nuclear size control in fission yeast. *The Journal of cell biology*, 179(4):593–600, 2007.