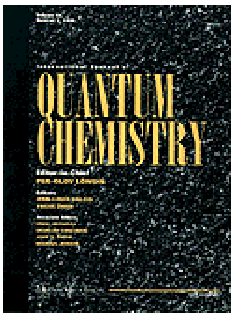

Online ISSN: 1097-461X Print ISSN: 0020-7608
International Journal of Quantum Chemistry
Forthcoming Articles Pages: NA

Published Online: 8 Aug 2001

Calculating populations of subcellular compartments using density matrix formalism

Vadim Alexandrov [1 *], Mark Gerstein [1 2]

[1]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave., New Haven, Connecticut 06511

[2]Department of Computer Science, Yale University, 266 Whitney Ave., New Haven, Connecticut 06511

email: Vadim Alexandrov (vadim.alexandrov@yale.edu)

[*]Correspondence to Vadim Alexandrov, Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave., New Haven, Connecticut 06511

Keywords

bioinformatics; quantum; localization; density matrix

Abstract

In our earlier paper (Ref. [1]) we developed an integrated probabilistic system for predicting the subcellular localization of proteins and estimating the relative populations of the various compartments in yeast. To justify our formulas we show here that there is a one-to-one correspondence between our previous calculations and the prediction of a state of a many-particle quantum system. The equivalence between these two types of predictions can be easily established if one maps the probability of finding a particular protein in a certain subcellular compartment to the probability of measuring the corresponding quantum particle in one of the possible quantum states (the number of proteins being equal to the number of particles in the system, and the number of compartments being equal to the number of achievable quantum states). Once the sought correspondence is established, we can utilize a well-known formula from quantum statistical mechanics to calculate the overall occupation of a particular quantum state, associating the state with the corresponding subcellular compartment. In the present work we present the details of how we arrived at the formula for the compartment population, borrowing the tools from quantum statistical mechanics. © 2001 John Wiley & Sons, Inc. Int J Quantum Chem, 2001

Received: 13 March 2001; Accepted: 24 May 2001

Digital Object Identifier (DOI)

10.1002/qua.1704 About DOI

Article Text

Introduction

In every living cell there is a complex machinery that sorts newly synthesized proteins and sends them to their final locations (subcellular compartments). Without an accurate estimate of the compartmental populations, it is hard to comprehend how a particular type of cell functions as a whole at the molecular level. However, with the advent of whole-genome sequencing, we now know the sequences of many proteins without knowing their localization.

In our earlier work [1] we developed a way of estimating the overall compartment population (i.e., the total number of proteins in a compartment) without rigidly localizing proteins to a single compartment. For each protein n in the cell we constructed a discrete likelihood distribution that we called a *probability state vector*, $\mathbf{P}_n$, whose coordinates are the probabilities to find a particular protein in each of the subcellular compartments. In order to estimate the total protein population for a particular compartment, we proposed to extract the corresponding components from the probability state vectors and add them up. The closest integer to the result of this summation would be the sought occupation number (population) for the given compartment. This approach differs from those previously suggested [2-8] in that it estimates the overall population without requiring individual predictions for every individual protein. While this summation of probabilities may appear to be intuitively obvious, its formal justification is not trivial. Fortunately, the necessary mathematical formalism to accomplish a very similar task of estimating the state occupation probabilities (and, thus, overall populations) was developed and successfully applied a long time ago in quantum mechanics [9-11]. We recognized this simple fact and established a one-to-one correspondence between the probability state vectors and the set of predictions regarding the results of measurements performed on a hypothetical quantum mechanical system.

Theory

CALCULATION OF STATE POPULATIONS IN STATISTICAL QUANTUM MECHANICS

In general, the incomplete information one has about the system usually presents itself in statistical quantum mechanics in the following way: the state of this system may be either the state $|\Psi_1\rangle$ with a probability p_1 or the state $|\Psi_2\rangle$ with a probability p_2, etc., such that the set of these probabilities p_j forms a distribution, i.e., $\Sigma_j p_j=1$. We then say that we are dealing with a *statistical mixture* of states $|\Psi_1\rangle,|\Psi_2\rangle$,

$\ldots$ with probabilities $p_1, p_2, \ldots$ (i.e., the initial state vector of the system is not perfectly known). Note that these states need not necessarily be eigenstates of any operator, or be orthogonal, but they can always be chosen normalized.

In case of the so-called *pure* state $|\Psi\rangle \equiv |\Psi_l\rangle$ (when all probabilities p_j are zero except $p_{j=l}$), the probability $P(a_n)$ to obtain a measurement a_n of an observable **A** can be found simply from the fourth postulate of quantum mechanics:

$$P(a_n) \equiv P_l(a_n) = \langle \psi_l | \hat{\mathbf{Q}}_n | \psi_l \rangle, \qquad 1$$

where $\hat{\mathbf{Q}}_n$ is a projector onto the eigensubspace associated with a_n. Introducing the density operator $\hat{\rho} \equiv |\psi_l\rangle\langle\psi_l|$ for the pure state $|\Psi_l\rangle$, the above expression can also be written as

$$P(a_n) = \frac{\mathrm{Tr}\{\hat{\mathbf{Q}}_n \hat{\rho}\}}{\mathrm{Tr}\{\hat{\rho}\}}. \qquad 2$$

In case of the statistical mixture of states, the probability of system to be found in state $|\Psi_l\rangle$ has probability p_l, therefore it is clear that the result in (1) must be weighted by p_l and summed over all values of l, i.e.,

$$P(a_n) = \sum_l P_l(a_n) = \sum_l p_l \frac{\mathrm{Tr}\{\hat{\mathbf{Q}}_n \hat{\rho}_l\}}{\mathrm{Tr}\{\hat{\rho}_l\}} = \frac{\mathrm{Tr}\{\hat{\mathbf{Q}}_n \hat{\rho}\}}{\mathrm{Tr}\{\hat{\rho}\}}, \qquad 3$$

the formula widely used in quantum statistical mechanics. In the above equation we also introduced the density operator $\hat{\rho}$ for the whole system,

$$\hat{\rho} \equiv \sum_l p_l \hat{\rho}_l = \sum_l p_l |\psi_l\rangle\langle\psi_l|. \qquad 4$$

Let us mention here that if the states $|\Psi_l\rangle$ are orthogonal, then the density operator $\hat{\rho}$ in this representation will have a diagonal form.

OUR APPROACH: BUILDING THE BRIDGE

Now, we notice that the state vector for a particular protein n (total number of proteins equal to N),

$$\mathbf{P}_n \equiv \left(p_n(\mathrm{loc}\,1), p_n(\mathrm{loc}\,2), \ldots, p_n(\mathrm{loc}\,M)\right), \qquad \sum_{\mathrm{loc}} p_n(\mathrm{loc}) = 1, \qquad 5$$

representing the probabilities $p_n(l)$ of this protein to be found in one of the M subcellular compartments, *can be thought of as a set of predictions concerning the results of measurements performed on a quantum mechanical system consisting of N distinguishable particles, each of which can be found in one of the M states*. Since the subcellular compartments do not overlap (one particular protein can not simultaneously belong to two different compartments), we can claim that the possible states of the introduced hypothetical quantum mechanical system are orthogonal. Therefore, the density matrix corresponding to each individual protein will necessarily have the diagonal form

$$\rho_n = \begin{pmatrix} p_n(1) & 0 & \dots & 0 \\ 0 & p_n(2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n(M) \end{pmatrix}. \qquad 6$$

Note that Tr{ρ_n}=1 for each n, manifesting the conservation of probability for each particle.

The projector onto a particular state subspace (corresponding to a particular subcellular compartment, for example, the first one) then assumes the following matrix form:

$$\mathbf{q}_n(1) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \qquad 7$$

One can see how this projector onto the subspace spanned by the first state (compartment) effectively extracts the probability for particle n (protein n) to be found in state one (compartment one):

$$\begin{aligned} \mathrm{Tr}\{\mathbf{q}_n(1)\rho_n\} = \mathrm{Tr}\left\{ \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \right. \\ \left. \times \begin{pmatrix} p_n(1) & 0 & \dots & 0 \\ 0 & p_n(2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_n(M) \end{pmatrix} \right\} \\ = \mathrm{Tr}\left\{ \begin{pmatrix} p_n(1) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \right\} \\ = p_n(1) \end{aligned} \qquad 8$$

The projector onto this subspace for the whole system of N distinguishable particles will be just a direct sum (denoted by $\oplus$ below) of projectors onto this subspace summed over all particles,

$$\hat{\mathbf{Q}}_1 \equiv \sum_{n=1}^{N} \oplus \hat{\mathbf{q}}_n(1), \qquad 9$$

or, in the matrix notation,

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 0 & 0 \ldots & & & & & \ldots & 0 \\ 0 & 0 & 0 & & & & & & \ldots \\ 0 & 0 & 0 & & & & & & \\ \ldots & & & 1 & 0 & 0 & & & \\ & & & 0 & 0 & 0 & & & \\ & & & 0 & 0 & 0 & & & \\ & & & & & & \ldots & & \\ & & & & & & \ldots & & \\ & & & & & & \ldots & & \\ & & & & & & \ldots & & \ldots \\ & & & & & & & 1\,0 & 0 \\ & & & & & & & 0\,0 & 0 \\ \ldots & & & & & & & & \\ 0 & \ldots & & & & & & \ldots 0\,0 & 0 \end{pmatrix}.$$

$$\hat{\rho} \equiv \sum_{n=1}^{N} \oplus \hat{\rho}_n, \qquad 10$$

or, in the matrix form,

$$\rho = \begin{pmatrix} p_1(1) & 0 & 0 & \ldots & & & & & & \ldots & 0 \\ 0 & p_1(2) & 0 & & & & & & & & \ldots \\ & & & & & & & & & & \\ 0 & 0 & p_1(M) & & & & & & & & \\ \ldots & & & p_2(1) & 0 & 0 & & & & & \\ & & & 0 & p_2(2) & 0 & & & & & \\ & & & & & & & & & & \\ & & & 0 & 0 & p_2(M) & & & & . & \\ & & & & & & \ldots & & & & \\ & & & & & & \ldots & & & & \\ & & & & & & \ldots & & & & \\ & & & & & & \ldots & & & & \ldots \\ & & & & & & & p_N(1) & 0 & & 0 \\ & & & & & & & 0 & p_N(2) & & 0 \\ \ldots & & & & & & & & & & \\ 0 & \ldots & & & & & \ldots & 0 & 0 & & p_N(M) \end{pmatrix}$$

$$n_1 = \frac{\mathrm{Tr}\{\mathbf{Q}_1 \rho\}}{\mathrm{Tr}\{\rho\}}. \qquad 11$$

The corresponding occupation number (population) for this state will be just

$$\begin{aligned} N_1 &= N \cdot n_1 = N \cdot \frac{\mathrm{Tr}\{\mathbf{Q}_1 \rho\}}{\mathrm{Tr}\{\rho\}} = N \cdot \frac{\mathrm{Tr}\{\mathbf{Q}_1 \rho\}}{N \cdot \mathrm{Tr}\{\rho_1\}} \\ &= N \cdot \frac{\mathrm{Tr}\{\mathbf{Q}_1 \rho\}}{N \cdot 1} = \mathrm{Tr}\{\mathbf{Q}_1 \rho\} = \sum_{n=1}^{N} p_n(1). \end{aligned} \qquad 12$$

What might seem to be natural even without using the density matrix formalism, i.e., summing over the probabilities related only to the first compartment in (12) to get the total occupancy for this compartment, is now rigorously proved to be a valid mathematical procedure justifying our heuristic protein localization approach described earlier.

One can easily think of the further extensions of the discussed formalism to study protein bulk traffic among compartments. For instance, if we assume that proteins in the cell are not necessarily confined to a particular compartment (e.g., a particular protein may contain several different localization signal sequences), the quantum states of the corresponding quantum system will not be orthogonal any more, introducing the nondiagonal elements in the density matrix. Therefore, by diagonalizing the density matrix it should be possible to find the "natural" basis vectors of intercompartmental traffic, which may provide further insight for a biologist about the mysterious dynamics of the cell functions.

Conclusion

By establishing a one-to-one correspondence of the original bioinformatics problem with a well-known formula for the occupation probability in statistical quantum mechanics we justified our previously proposed method for calculating populations of subcellular compartments. Being intrinsically probabilistic by nature, quantum theory turns out to be useful in contemplating the nascent statistical laws of bioinformatics and providing a new perspective for exploration of this emerging field.

References

1 Drawid, A.; Gerstein, M. *J Mol Biol* 2000, **301**, 1059-1075.

2 Nakai, K.; Kanehisa, M. *Prot Struct Funct Genet* 1991, **11**, 95-110.

3 Nakai, K.; Kanehisa, M. *Genomics* 1992, **14**, 897-911.

4 Nakai, K.; Horton, P. *Trends Biochem Sci* 1999, **24**, 34-36.

5 Nakai, K.; Horton, P. *Intelligent Systems for Molecular Biology* 1996, **4**, 109-115.

6 Nakai, K.; Horton, P. *Intelligent Systems for Molecular Biology* 1997, **5**, 147-152.

7 Reinhardt, A.; Hubbard, T. *Nucleic Acids Res* 1998, **26**, 2230-2236.

8 Andrade, M.; O'Donoghue, S.; Rost, B. *J Mol Biol* 1998, **276**, 517-525.

9 Bohm, A. in *Quantum Mechanics: Foundations and Applications*, 2nd ed.; Springer: New York, 1986.

10 Cohen-Tannoudji, C.; Diu, B.; Laloe, F. in *Méchanique Quantuque*, 2nd ed.; Hermann: Paris, 1977.

11 Landau, D.; Lifshitz, E. in *Kvantovaia Mehanika* (*Course of Theoretical Physics*; Vol. **3**); Nauka: Moscow, 1980.
