Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 5345–5349, August 1988
Applied Mathematical Sciences

Discrete-time random walks on diagrams (graphs) with cycles

(directional diagrams/state probabilities/cycle fluxes/tennis examples)

TERRELL L. HILL

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Terrell L. Hill, March 23, 1988

ABSTRACT **After a review of the diagram method for continuous-time random walks on graphs with cycles, the method is extended to discrete-time random walks. The basic theorems carry over formally from continuous time to discrete time. Three problems in tennis probabilities are used to illustrate random walks on discrete-time diagrams with cycles.**

Properties and applications of diagrams with cycles have been summarized in a book (1) and have been extended considerably in three recent papers (2–4). All of this work relates to continuous-time random walks on the diagrams of interest. The purpose of the present paper is to extend the discussion to corresponding discrete-time problems. However, I begin with a summary of continuous-time properties because these are used as the starting point for the discrete-time discussion. Some of the notation here is a little different from that used in refs. 2–4.

Steady-State Properties of Continuous-Time Diagrams (Graphs) with Cycles

For concreteness, this review (1) is related in large part to a particular example, shown in Fig. 1. Suppose that a protein complex can exist in any one of $n = 5$ significant states, represented simply by the numbers (vertices) in Fig. 1*A*. Certain pairs of these states are interconvertible, as indicated by the lines (edges) in the state diagram (graph). The diagram has three cycles, a, b, and c, shown in Fig. 1*B*. The complex can be viewed as undergoing a continuous-time random walk from state to state along the lines of the diagram, Fig. 1*A*. The transition probabilities or rate constants that govern this walk are denoted α_{ij} (one for each direction along each line, though some of these may be negligibly small). For example, $\alpha_{12}dt$ is the probability that a complex in state 1 makes a transition to state 2 in the infinitesimal time interval dt. The complex is always in some one state: the transitions themselves are instantaneous. The α_{ij} are given at the outset as part of the diagram and are independent of t. We imagine following the random walk among the states of the diagram for a very long time. The two questions we consider are (*i*) what fraction of time p_i is spent by the complex in each state i? (*ii*) at what mean rates J_{a+}, J_{a-}, etc., are the three cycles completed in each direction (+ or − in Fig. 1*B*)? The cycle completions are of particular interest because they tell what the complex accomplishes in the course of its long random walk.

For *any* arbitrary finite diagram (graph), with all α_{ij} given, the same questions can be asked about its n stationary-state probabilities p_i and its mean rates of cycle completions $J_{\kappa\pm}$.

There are two theorems (pages 6–10, 17–22, and 201–205 of ref. 1) that allow one to find the p_i and the $J_{\kappa\pm}$ by a graphical procedure from the given α_{ij} for an arbitrary finite diagram (graph) with cycles. The example in Fig. 1 is used to illustrate the two procedures or algorithms.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

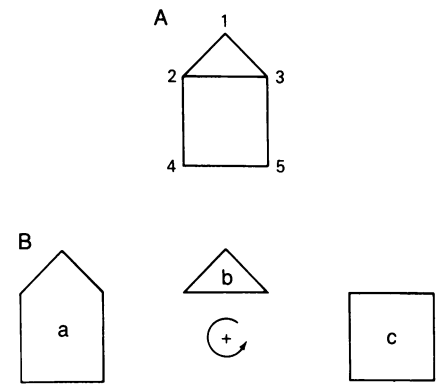

FIG. 1. (*A*) Kinetic diagram (or graph) showing the $n = 5$ significant states of a hypothetical protein complex and possible transitions between some pairs of states. (*B*) Cycles belonging to the diagram. The choice of positive direction is arbitrary.

It should be mentioned first that the stationary p_i can be found by Cramer's rule from the linear equations (Fig. 1)

$$p_1 + p_2 + p_3 + p_4 + p_5 = 1 \qquad [1]$$

and any four kinetic equations such as (see Fig. 1*A*)

$$\frac{dp_1}{dt} = 0 = (\alpha_{21}p_2 - \alpha_{12}p_1) + (\alpha_{31}p_3 - \alpha_{13}p_1). \qquad [2]$$

A graphical algorithm (1, 5, 6) provides an alternative solution for the p_i that is more interesting and elegant mathematically.

The first step is to construct from the diagram the complete set of *partial diagrams* (subgraphs), each of which contains the maximum possible number of lines ($n - 1$) that can be included without forming any cycles (these subgraphs are trees). There are 11 partial diagrams for Fig. 1*A*. The next step is to introduce arrows into each line of the partial diagrams in five different ways, one way for each state. The figures thus generated (there are 55) are called *directional diagrams*, because of the arrows. For example, for state 2, all arrows "flow" toward state 2, possibly by convergent "streams" but never by divergent streams. The same is true for the directional diagrams belonging to each of the other states.

Each of the 55 directional diagrams has an algebraic value that is obtained by forming the product of the transition probabilities α_{ij} indicated by each of its $n - 1$ arrows. Let Σ be the sum of the algebraic values of all 55 directional diagrams. Then the final statement of the theorem is

$$p_i = (\text{sum of algebraic values of directional diagrams belonging to state } i)/\Sigma. \qquad [3]$$

The generalization of Eq. 3 to an arbitrary diagram is obvious. If the diagram has many more states than in this example, the graphical method has aesthetic appeal only; for practical results one would have to solve the linear algebraic equations, analogues of Eqs. **1** and **2**, numerically by computer.

Eq. **3** is intuitively plausible because the flow of arrows toward a state is associated with transition probabilities that lead toward the state and increase its chance of occupation. A similar flow toward cycles appears in the second algorithm (1, 6), which we now turn to.

We start, in our example, with the set of 11 partial diagrams. These subgraphs have two lines (edges) missing and no cycles. We now add one line, in two ways, to each partial diagram. These (there are 22) are also subgraphs, but now each contains one cycle and n lines. Of the 22 subgraphs generated, only six are different. In each of the six cases we now add arrows to all lines that are appendages to the cycles: the arrows are inserted so that they flow toward the cycles; streams of arrows may converge but not diverge, just as above for directional diagrams. Thus we are led, in this example, to six *flux diagrams*, classified according to cycle. Σ_κ is a property of cycle κ and is the sum of α_{ij} products for the appendages of the flux diagrams that belong to cycle κ. This is illustrated in Fig. 2 for cycles a, b, and c.

Finally, we define $\Pi_{\kappa+}$ as the product of the α_{ij} around cycle κ in the + direction, and similarly for $\Pi_{\kappa-}$. In the example,

$$\Pi_{a+} = \alpha_{12}\alpha_{24}\alpha_{45}\alpha_{53}\alpha_{31}, \qquad \Pi_{b-} = \alpha_{13}\alpha_{32}\alpha_{21}, \qquad [4]$$

etc. Then the theorem relating to cycle fluxes (mean rates of cycle completions) is, for any cycle κ,

$$J_{\kappa+} = \Pi_{\kappa+}\,\Sigma_\kappa/\Sigma, \qquad J_{\kappa-} = \Pi_{\kappa-}\,\Sigma_\kappa/\Sigma. \qquad [5]$$

The net mean rate of κ cycle completions in the + direction is

$$J_\kappa = J_{\kappa+} - J_{\kappa-}. \qquad [6]$$

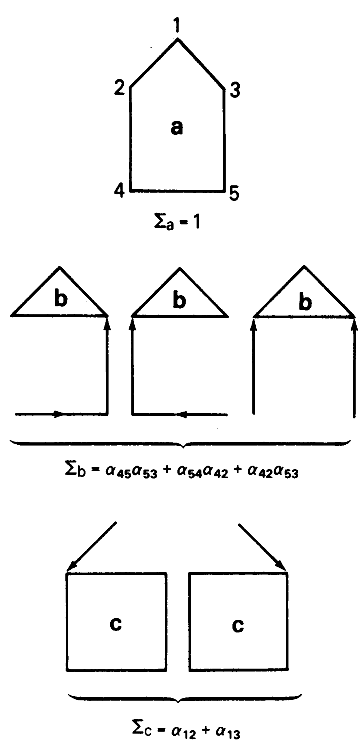

FIG. 2. The six flux diagrams belonging to the diagram in Fig. 1A, classified according to cycle.

Note also that

$$J_{\kappa+}/J_{\kappa-} = \Pi_{\kappa+}/\Pi_{\kappa-}. \qquad [7]$$

As an explicit example, from Fig. 2,

$$J_c = (\alpha_{24}\alpha_{45}\alpha_{53}\alpha_{32} - \alpha_{23}\alpha_{35}\alpha_{54}\alpha_{42})(\alpha_{12} + \alpha_{13})/\Sigma, \qquad [8]$$

where Σ contains 55 terms (see above).

At this point I digress to add that I gave (1, 6) a correct proof of Eq. **6** but an incorrect "proof" of Eqs. **5** (page 22 of ref. 1). The latter deficiency has been removed by Kohler and Vollmerhaus (7) and by Qian *et al.* (8, 9).

Each term in the numerators of Eqs. **5** is a product of n factors α_{ij}. The terms in the denominator (Σ) have $n - 1$ such factors.

In complicated cases, $J_{\kappa+}$ and $J_{\kappa-}$ can be calculated exactly by a linear algebraic method (2), as an alternative to Eqs. **5**. Ref. 2 also includes a different and simple proof of Eqs. **5**.

The *transition flux* J_{ij} between states i and j is defined as

$$J_{ij} = \alpha_{ij}p_i - \alpha_{ji}p_j. \qquad [9]$$

In the course of the proof of Eqs. **3** and **6**, one finds that J_{ij} is equal to the algebraic sum of cycle fluxes for those cycles that include the line ij. A plus sign is used for a term in the sum if the cycle + direction coincides with the direction $i \rightarrow j$; otherwise a minus sign is used. For example, in Fig. 1,

$$J_{12} = J_a + J_b, \qquad J_{23} = J_b - J_c. \qquad [10]$$

Let α_i be the sum of all transition probabilities *out of* state i. For example, in Fig. 1A, $\alpha_2 = \alpha_{21} + \alpha_{23} + \alpha_{24}$. If the system (random walker, complex) is in state i at, say, $t = 0$, the probability that any transition out of state i first occurs between t and $t + dt$ is $\alpha_i e^{-\alpha_i t}dt$. The mean time at which this transition occurs is then $\bar{t} = 1/\alpha_i$. That is, $1/\alpha_i$ is the mean lifetime of state i whenever it is reached in the random walk. When the transition does occur, the probability that the transition is from state i to a particular state j is α_{ij}/α_i. The above comments are needed below.

Normalized Transition Probabilities

This section relates to a formal modification of the problem in the previous section. The modification is to replace every transition probability α_{ij} in the diagram of interest by the dimensionless quantity α_{ij}/α_i. This is still a continuous-time two-way random walk. The new outgoing transition probabilities for every state i are normalized to unity. Hence the mean lifetime in any state before a transition occurs is unity. The theorems, Eqs. **3** and **5**, can now be applied to the new set of normalized transition probabilities. When this is done, let us use the new notation P_i for state probabilities, σ in place of Σ, and $j_{\kappa\pm}$ for the cycle fluxes. What is the relation between p_i and P_i and between $J_{\kappa\pm}$ and $j_{\kappa\pm}$?

Define the product $\Pi \equiv \alpha_1\alpha_2 \ldots \alpha_n$. The numerators in Eqs. **5** contain, in each term, a product with one outgoing rate constant from each of the n states. Hence,

$$j_{\kappa+} = \frac{(\Pi_{\kappa+}\Sigma_\kappa/\Pi)}{\sigma}; \text{ etc.} \qquad [11]$$

Eqs. **5** and **11** have $\Pi_{\kappa+}\Sigma_\kappa$ in common, so we find on elimination of $\Pi_{\kappa+}\Sigma_\kappa$,

$$j_{\kappa+} = \left(\frac{\Sigma}{\Pi\sigma}\right)J_{\kappa+}, \text{ etc.} \qquad [12]$$

Note that $j_{\kappa+}$ is dimensionless. Because Σ, Π, and σ are all independent of the choice of cycle κ, $j_{\kappa+}$ and $J_{\kappa+}$ differ only by a constant normalization factor. Therefore, the *relative* frequency of completions of the different possible cycles is the same in the two cases (i.e., whether the transition probabilities are α_{ij} or α_{ij}/α_i). This is intuitively obvious because the *sequence* of states and cycle completions in the two random walks would not be altered when $\alpha_{ij} \rightarrow \alpha_{ij}/\alpha_i$. This follows because the *sequence* in the α_{ij} case is actually determined by the probabilities α_{ij}/α_i.

Let (sum) denote the numerator in Eq. **3**. Each term in (sum) contains an outgoing transition probability from every state except i. Hence, in the modified problem,

$$(\text{sum}) \rightarrow (\text{sum})\alpha_i/\Pi.$$

Thus,

$$P_i = \frac{(\text{sum})\alpha_i/\Pi}{\sigma}. \qquad [13]$$

Then, from Eq. **3**, on eliminating (sum),

$$P_i = \left(\frac{\Sigma}{\Pi\sigma}\right) p_i \alpha_i. \qquad [14]$$

Aside from the leading normalization factor, this result is also to be expected intuitively. That is, p_i in the actual random walk (i.e., in the α_{ij} case) is proportional to the product of the relative frequency P_i that state i is visited in the course of the long random walk and to the mean time $1/\alpha_i$ spent in state i on each visit.

Although the connection between Σ and σ is straightforward, they are not related mathematically in a really simple way. However, summation of Eq. **14** over i gives

$$\frac{\Sigma}{\Pi\sigma} = \left(\sum_{i=1}^{n} p_i \alpha_i\right)^{-1} = \sum_{i=1}^{n} (P_i/\alpha_i). \qquad [15]$$

Thus, physically, $\Sigma/\Pi\sigma$ is the overall mean time between transitions in the actual random walk.

Discrete-Time Random Walks

Normalized transition probabilities can be used in Eqs. **3** and **5** to find the P_i and $j_{\kappa\pm}$, as just described. The mean lifetime in any state is unity, but there is a distribution in lifetimes, with probability $e^{-t}dt$ for a lifetime between t and $t + dt$. That is, this is a continuous-time problem. Consider now a modification of this system in which the lifetime distribution above is replaced by a δ function at $t = 1$. This modification will have no effect on P_i because, in a very long random walk, the fraction of the total time spent in state i will obviously be the same whether the lifetime distribution is e^{-t} (with mean at $t = 1$) or is a δ function at $t = 1$. Similarly, $j_{\kappa+}$, etc., will be unchanged by the modification because the mean rate of completing any cycle depends on the sequence of states in the long random walk and on the mean time spent in each state on each visit. In the modification, the sequence is identical and the mean time is unaltered.

In summary, Eqs. **3** and **5** can still be applied when the outgoing transition probabilities from any state are normalized to unity and when the time between transitions is always unity. But this is just a conventional discrete-time random walk in which the time is counted simply by counting transitions or steps. The unconventional part of the problem is that the random walk occurs on a finite graph with cycles.

The notation we use for Eqs. **3** and **5** in this special case, with normalized outgoing transition probabilities, is

$$P_i = (\text{sum of algebraic values of directional diagrams belonging to state } i)/\sigma \qquad [16]$$

$$j_{\kappa+} = \pi_{\kappa+}\sigma_\kappa/\sigma, \qquad j_{\kappa-} = \pi_{\kappa-}\sigma_\kappa/\sigma. \qquad [17]$$

These relations are formally the same as Eqs. **3** and **5**; the notation change simply reminds one of the discrete-time special case. Of course the sum in Eq. **16** is not the same as that in Eq. **3** because of the use in Eq. **16** of normalized outgoing transition probabilities.

All of the quantities appearing in Eqs. **16** and **17** are dimensionless. For example, $j_{\kappa+}$ is the mean number of $\kappa+$ cycles completed per transition or step in the walk.

Tennis Problems as Examples of Discrete-Time Random Walks

In this section I use three tennis problems to illustrate discrete-time random walks on diagrams with cycles. Actually, each of these problems is originally an absorption problem on a diagram without cycles but when the original diagram is *closed*, as described in ref. 2, one-way cycles are produced. We work here with the closed diagrams, as examples of discrete-time diagrams with cycles. The method is (presumably) new, not the results. Because these cycles are one-way, it is possible to calculate cycle fluxes directly from state probabilities (this is not possible for arbitrary cycles).

Tennis Game That Starts at Deuce. As a very simple example, consider a game in which the winner is the first player to win two or more points with a margin of two points. Player 1(2) has a probability $p(q)$ of winning each point, where $p + q = 1$. To obtain stationary results, we imagine that the game is replayed very many times. That is, there is a very long discrete-time random walk on the closed diagram (graph) in Fig. 3 *A* and *B*: each transition or step is a point and completion of a game corresponds to completion of a cycle that returns the walk to state 0. The states (scores) are 0 = deuce; 10 = advantage player 1; and 01 = advantage player 2. The diagram has two cycles (Fig. 3*C*), labeled 1 and 2 (for the winner of the game).

The steady-state equations for the state probabilities are, from Fig. 3*B*,

$$P_0 = P_{01} + P_{10}, \qquad P_{01} = qP_0, \qquad P_{10} = pP_0. \qquad [18]$$

The solution (normalized) is

$$P_0 = 1/2, \qquad P_{10} = p/2, \qquad P_{01} = q/2. \qquad [19]$$

These are the probabilities of occurrence of the different possible scores when the game is repeated many times. The cycle fluxes are

$$j_1 = pP_{10} = p^2/2, \qquad j_2 = qP_{01} = q^2/2. \qquad [20]$$

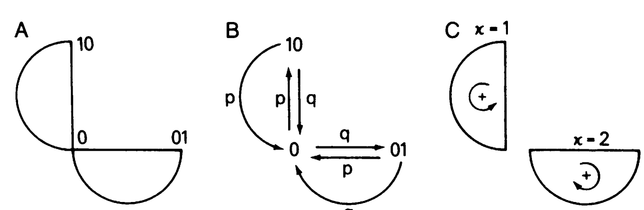

FIG. 3. (*A*) Diagram (graph) for a tennis game that starts at deuce. (*B*) Normalized (outgoing) transition probabilities in the diagram. (*C*) The two cycles of the diagram.

These are the mean number of games won by each of the two players per point played. The fraction of games won by player 1 is $p^2/(p^2 + q^2)$. The mean number of completed games per point is $j_1 + j_2$ or $(p^2 + q^2)/2$. The mean number of points per game is $2/(p^2 + q^2)$. It is left as an exercise for the interested reader to derive Eqs. **19** and **20** from Eqs. **16** and **17**.

Two-Out-of-Three Set Match. Player 1(2) has a probability $p(q)$ of winning each set, with $p + q = 1$. The closed diagram, including transition probabilities, is given in Fig. 4*A*. Unlike Fig. 3*B*, these are all one-way transitions. Here a transition or step in the walk corresponds to a set played. The state notation indicates the score in sets. State 0 represents the score 0–0. The match is repeated many times; completion of a match returns the state (completes a cycle) to 0. Fig. 4*B* displays the six cycles. In every case $j_{\kappa-} = 0$ and $j_\kappa = j_{\kappa+}$. The number of matches won by each player, per set played, is

$$j_1 = j_{1a} + j_{1b} + j_{1c}, \qquad j_2 = j_{2a} + j_{2b} + j_{2c}. \qquad [21]$$

The steady-state equations for the state probabilities are

$$P_0 = p(P_{10} + P_{11}) + q(P_{01} + P_{11})$$
$$P_{10} = pP_0, \qquad P_{01} = qP_0, \qquad P_{11} = qP_{10} + pP_{01}. \qquad [22]$$

The solution is

$$P_0 = 1/\sigma, \qquad P_{10} = p/\sigma, \qquad P_{01} = q/\sigma$$
$$P_{11} = 2pq/\sigma, \qquad \sigma = 2(1 + pq). \qquad [23]$$

The fluxes are

$$j_{1a} = pP_{10} = p^2/\sigma, \qquad j_{1b} + j_{1c} = pP_{11} = 2p^2q/\sigma$$
$$j_{2a} = qP_{01} = q^2/\sigma, \qquad j_{2b} + j_{2c} = qP_{11} = 2pq^2/\sigma \qquad [24]$$

$$j_1 = p^2(1 + 2q)/\sigma, \qquad j_2 = q^2(1 + 2p)/\sigma. \qquad [25]$$

Then the mean number of matches per set is

$$j_1 + j_2 = 1/\sigma = P_0 \qquad [26]$$

(flow into state 0 = flow out of state 0). The mean number of sets per match is σ. The fraction of matches won by player 1 is

$$j_1/(j_1 + j_2) = p^2(1 + 2q). \qquad [27]$$

The reader may wish to derive Eqs. **23** and **24** from Eqs. **16** and **17**.

Regular Tennis Game. As a final example, consider a regular tennis game: to win requires four or more points with a margin of at least two. Let $p(q)$ be the probability that player 1(2) wins each point, with $p + q = 1$. The closed diagram,

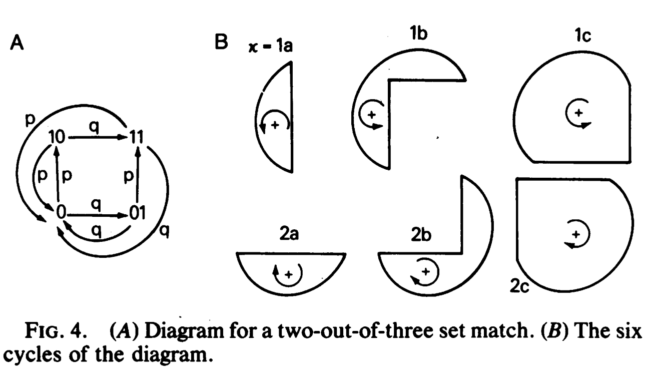

FIG. 4. (*A*) Diagram for a two-out-of-three set match. (*B*) The six cycles of the diagram.

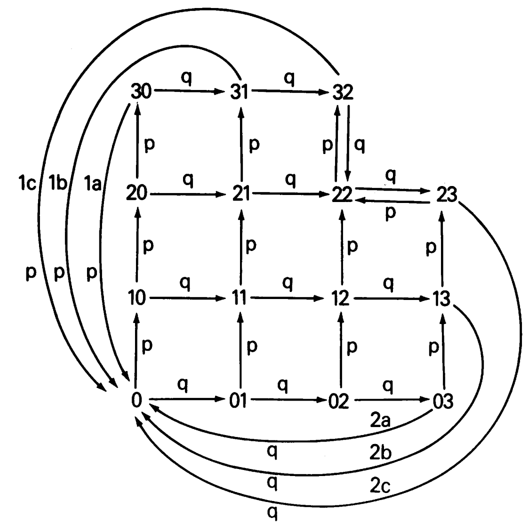

FIG. 5. Diagram for a regular tennis game. The labels 1a, 1b, etc., refer to game-completion transition fluxes, not to cycle fluxes.

with transition probabilities, is shown in Fig. 5. This is a generalization of Fig. 3*B*. Each transition or step in the random walk corresponds to a point played. The states in the diagram are point scores, with 32 representing any of 32, 43, 54, . . .; 22 representing any of 22, 33, 44, . . .; and 23 representing any of 23, 34, 45, ... In tennis language 32 is 40–30 or advantage player 1, 22 is 30–30 or deuce, etc. We imagine a very long random walk on this diagram, where the game (same server) is repeated many times. The labels 1a, 2a, etc., refer to game-completion transition fluxes, j_{1a}, etc., not to cycle fluxes. The rates at which the two players win games, per point played, are

$$j_1 = j_{1a} + j_{1b} + j_{1c} = p(P_{30} + P_{31} + P_{32}) \qquad [28]$$
$$j_2 = j_{2a} + j_{2b} + j_{2c} = q(P_{03} + P_{13} + P_{23}). \qquad [29]$$

Corresponding to Eqs. **1** and **2**, we have here

$$P_0 + \ldots + P_{32} + P_{22} + P_{23} = 1$$
$$P_{10} = pP_0, \qquad P_{01} = qP_0$$
$$P_{11} = qP_{10} + pP_{01}, \text{ etc.} \qquad [30]$$

These 15 linear equations in 15 unknowns (P_0, P_{ij}) are easy to solve. The results are

$$P_0 = N/D, \qquad P_{10} = pN/D, \qquad P_{20} = p^2N/D$$
$$P_{11} = 2pqN/D, \qquad P_{30} = p^3N/D, \qquad P_{21} = 3p^2qN/D$$
$$P_{31} = 4p^3qN/D, \qquad P_{32} = 10p^3q^2/D$$
$$P_{22} = 2p^2q^2(3 + 4pq)/D$$
$$N = p^2 + q^2, \qquad D = 4\,(1 - pq + 6p^3q^3). \qquad [31]$$

To obtain P_{ji} from P_{ij}, exchange p and q.

The total mean number of games played per point, as in Eq. **26**, is

$$j = j_1 + j_2 = P_0 = N/D. \qquad [32]$$

Thus the mean number of points per game is D/N. The fraction of games won by player 1 is

$$\frac{j_1}{j} = p^4 + 4p^4q + \frac{10p^4q^2}{p^2 + q^2}. \qquad [33]$$

Because of the complexity of Fig. 5, it would not be practical in this case to use Eqs. **16** and **17**.

1. Hill, T. L. (1977) *Free Energy Transduction in Biology* (Academic, New York).
2. Hill, T. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2879–2883.
3. Hill, T. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3271–3275.
4. Hill, T. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4577–4581.
5. King, E. L. & Altman, C. (1956) *J. Phys. Chem.* **60,** 1375–1378.
6. Hill, T. L. (1966) *J. Theor. Biol.* **10,** 442–459.
7. Kohler, H.-H. & Vollmerhaus, E. (1980) *J. Math. Biol.* **9,** 275–290.
8. Qian, C., Qian, M. & Qian, M. (1981) *Sci. Sin.* **24,** 1431–1448.
9. Qian, M., Qian, M. & Qian, C. (1984) *Sci. Sin.* **27,** 470–481.