

AN ALGORITHM TO ASSEMBLE PATHWAYS FROM PROCESSES

J. E. MITTENTHAL

Cell and Structural Biology, University of Illinois, 505 S. Goodwin St., Urbana, IL 61801

To understand or to modify a biological pathway, the first step is to determine the patterns of coupling among its processes that are compatible with its input-output relation. Algorithms for this purpose have been devised for metabolic pathways, in which the reactions typically leave the enzymes unmodified. As shown here, one of these algorithms can also assemble molecular networks in which reactions modify proteins, if the proteins are included among the inputs to the reactions. Thus one procedure suffices to assemble pathways for metabolism, cytoplasmic signal transduction, and gene regulation.

1 Introduction

In a molecular pathway or network processes are coupled, with outputs of some processes used as inputs to others. We can discern the coupling and assemble processes into pathways intuitively, for small sets of processes that assemble in a unique way. However, as the number of processes grows, they may form several pathways that can mediate a given input-output relation. Then tools that aid the assembly can be useful. With these tools we can generate an ensemble of alternative pathways, simulate their dynamics, and compare their performance. If knowledge of the real pathway is fragmentary, such an inquiry may point out gaps in our understanding and so may channel further research. If the real pathway is well characterized, study of an ensemble of alternatives can show whether it is an optimum of the evolutionary search (Melendez-Hevia et al., 1994; Mittenthal et al., 1993, 1996). To modify a pathway through bioengineering it is necessary to explore alternatives that more nearly approach a desired performance (Bailey, 1991).

Various algorithms have been used to assemble metabolic reactions into pathways (Serio and Bailey, 1988; Happel et al., 1990; Mavrouniotis et al., 1990, 1992). Typically in metabolism reactions do not modify enzymes, so the enzymes do not appear in the algorithms for constructing pathways. In other networks a process may provide substrates, cofactors, and regulators, as in metabolism, but may also modify proteins. For example, in the signaling net for cytoplasmic signal transduction, reactions modify enzymes and regulatory proteins. Algorithms to synthesize such networks are needed but are not yet available. As

illustrated here, the algorithm that Mavrovouniotis et al. (1990) developed for constructing metabolic networks can also be used to construct protein-modifying networks, by including the proteins in the reactions. This algorithm, unlike the others mentioned above, does not depend on stoichiometry, can deal with irreversible reactions, and is complete in that it provides all pathways that meet the constraints.

2 Reactions

The method uses a list of reactions that may participate in the network. The list may include hypothetical reactions, or reactions to be excluded from a particular assembly. A reaction is a unidirectional transformation; the forward and reverse directions of a bidirectional transformation are treated as two reactions. The statement of the reaction includes the reactants and their stoichiometry, the enzymes that catalyze the reaction, and cofactors or regulatory factors that are not transformed in the reaction.

In a reaction molecules may associate and change their conformations without making or breaking covalent bonds. For example, binding an enzyme to a regulatory protein may activate an inactive enzyme, or inactivate an active enzyme. Other reactions may make or break covalent bonds. Examples include phosphorylation of a protein by a kinase or its dephosphorylation by a phosphatase.

3 Steps of the method

Specify the constraints.

- What is the status of each molecule? Is it an admissible participant in the network? Which of its attributes are real, and which hypothetical?
- What are the roles allowed for each molecule? Can it be an input, output, and/or intermediate of the network?
- For each role, is the molecule required, allowed, or excluded?

Make the reaction list.

- For each molecule that may be an intermediate, check that at least one reaction produces it, and at least one reaction consumes it.

- Label the reactions. If the same molecules are involved together in more than one reaction, as in the forward and reverse directions of a bidirectional transformation, it may be useful to give these reactions the same primary index and different secondary indices.

Make the molecule list. For every molecule in the reaction list, enter each form of the molecule separately. Record the indices of all reactions in which it participates, the way it participates in each reaction -- as substrate, product, enzyme, or cofactor - - and the number of these reactions.

Reconstruct the network. Associate with each reaction the reactions that provide its substrates, catalysts, or cofactors. Eliminate each molecule from the molecule list after the reactions that produce it have been associated with the reactions that use it. To simplify recording, eliminate sooner the molecules involved in fewer reactions. A set of coupled reactions identified in this way is a partial pathway. The reactions themselves are the initial, shortest partial pathways. Iterating the association of reactions that produce a molecule with reactions that use it constructs partial pathways of increasing length. After this procedure augments a partial pathway, replace the previous version with the augmented version in the reaction list. When the iteration is complete, the partial pathways can be displayed as directed graphs, and their dynamics can be simulated.

4 Example: Regulation of the breakdown and synthesis of glycogen

We examine the network that regulates the availability of glucose in vertebrate skeletal muscle cells. Two intracellular second messengers, cyclic AMP (adenosine-3',5'-monophosphate) and calcium ion, have partially antagonistic and partially synergistic effects. In response to adrenalin, cyclic AMP stimulates the breakdown of glycogen by glycogen phosphorylase and inhibits the synthesis of glycogen by glycogen synthase. Stimulation of the muscle releases calcium, which transiently increases the activation of glycogen phosphorylase but soon reverses the response to cAMP. Hardie (1991) provided a description of this system that is the basis for the reaction list. Some reactions known to be relevant are not considered, for simplicity. These reactions include phosphorylation of glycogen synthase by casein kinase-1 and glycogen synthase kinase-3, and some regulatory interactions of protein phosphatase-1.

4.1 The molecule list and the reaction list

- Molecule list: This list introduces the names and abbreviations for molecules. Reactions of each molecule are taken from the reaction list that follows.
+ means that the molecule is synthesized through this reaction. - means it is degraded. c means the molecule is a catalyst or cofactor in this reaction.
(4) means that the molecule is involved in four reactions.

AMP: adenosine-5'-monophosphate	(1)
+: 8.1	
Ca: cytoplasmic calcium	(6)
+: 2.4; 7.2; 8.2 -: 2.3; 7.1; 8.1	
cAMP: cyclic adenosine-3',5'-monophosphate	(3)
+: 1.2 -: 1.1; 8.3	
PKA: cAMP-dependent protein kinase; protein kinase A; A-kinase	(5)
+: 1.1 -: 1.3 c: 2.1; 4.1; 5.2	
PKAR2: A-kinase, type 2 regulatory subunit	(4)
+: 1.2; 1.5 -: 1.3; 1.4	
PKAR2-P: A-kinase, type 2 regulatory subunit, phosphorylated	(2)
+: 1.4 -: 1.5	
cAMP,PKAR2: complex of cAMP and PKAR2	(2)
+: 1.1 -: 1.2	
PKAR2,PKA: complex of PKAR2 and PKA	(2)
+: 1.2 -: 1.1	
PlaseK: phosphorylase kinase	(2)
+: 2.2 -: 2.1	
PlaseK-P: phosphorylase kinase, phosphorylated	(4)
+: 2.1, 2.4 -: 2.2, 2.3 c: 6.1	
PlaseK-P,Ca: phosphorylase kinase, phosphorylated, Ca-activated	(3)
+: 2.3 -: 2.4 c: 6.1	
PP1: protein phosphatase-1	(5)
+: 3.1 -: 3.2 c: 2.2; 5.1; 6.2	
PP1I: protein phosphatase-1 inhibitor-1	(2)
+: 4.2 -: 4.1	
PP1I-P: protein phosphatase-1 inhibitor-1, phosphorylated	(4)
+: 3.1; 4.1 -: 3.2; 4.2	
PP1,PP1I-P: complex of PP1 and PP1I-P	(2)
+: 3.2 -: 3.1	
PP2B: protein phosphatase-2B	(2)
+: 7.2 -: 7.1	

PP2B,Ca: protein phosphatase-2B, Ca-activated	(4)
+: 7.1 -: 7.2 c: 1.5; 4.2	
GlySyn: glycogen synthase	(2)
+: 5.1 -: 5.2	
GlySyn-P: glycogen synthase, phosphorylated	(2)
+: 5.2 -: 5.1	
Plase: glycogen phosphorylase b.	(2)
+: 6.2 -: 6.1	
Plase-P: glycogen phosphorylase a	(2)
+: 6.1 -: 6.2	
cNMPde: cyclic nucleotide phosphodiesterase	(2)
+: 8.2 -: 8.1	
cNMPde,Ca: cyclic nucleotide phosphodiesterase, activated	(3)
+: 8.1 -: 8.2 c: 8.3	

- Reaction list: rxn# = reaction number (index). act = activate. inh = inhibit..

rxn#	effect	reaction	prerequisites
1.1	act PKA	cAMP + PKAR2,PKA -> cAMP,PKAR2 + PKA	---
1.2	pre-inh PKA	cAMP,PKAR2 -> cAMP + PKAR2	---
1.3	inh PKA	PKAR2 + PKA -> PKAR2,PKA	---
1.4	act PKA	PKAR2 -> PKAR2-P	PKAR2
1.5	inh 1.3	PKAR2-P -> PKAR2	PP2B,Ca
2.1	act PlaseK	PlaseK -> PlaseK-P	PKA
2.2	inh PlaseK	PlaseK-P -> PlaseK	PP1
2.3	act PlaseK	PlaseK-P + Ca -> PlaseK-P,Ca	high Ca
2.4	inh PlaseK	PlaseK-P,Ca -> PlaseK-P + Ca	low Ca
3.1	act PP1	PP1,PP1I-P -> PP1 + PP1I-P	---
3.2	inh PP1	PP1 + PP1I-P -> PP1,PP1I-P	---
4.1	act PP1I	PP1I -> PP1I-P	PKA
4.2	inh PP1I	PP1I-P -> PP1I	PP2B,Ca
5.1	act GlySyn	GlySyn-P -> GlySyn	PP1
5.2	inh GlySyn	GlySyn -> GlySyn-P	PKA
6.1	act Plase	Plase -> Plase-P PlaseK-P; PlaseK-P,Ca	
6.2	inh Plase	Plase-P -> Plase	PP1
7.1	act PP2B	PP2B + Ca -> PP2B,Ca	high Ca
7.2	inh PP2B	PP2B,Ca -> PP2B + Ca	low Ca
8.1	act cNMPde	cNMPde + Ca -> cNMPde,Ca	high Ca
8.2	inh cNMPde	cNMPde,Ca -> cNMPde + Ca	low Ca
8.3	inh cAMP	cAMP -> AMP	cNMPde,Ca

4.2 Construct partial pathways

If reaction r' provides a substrate for reaction r'' , show this by replacing r'' with (r', r'') in the reaction list. If reaction r provides a catalyst or cofactor for r'' , in the molecule list replace r'' with (cr, r'') . The convention for listing precursors is (cr, r', r'') . When this procedure has been followed for all reactions r'' for which a given molecule is a prerequisite, it can be eliminated from the molecule list. Eliminate first the molecules involved in fewest reactions.

To begin, note that AMP is involved in only one reaction but is not a prerequisite for any other reactions, so we simply eliminate it. PKAR2-P is a substrate for reaction 1.5, and it is produced in reaction 1.4 ; it is only involved in these two reactions. Therefore we replace reaction 1.5 by (1.4, 1.5) in the reaction list, and eliminate PKAR2-P from the molecule list. We proceed in an analogous way for other molecules. However, reactions that release Ca when the concentration of Ca is low are not used as sources of Ca for reactions that take up Ca when the concentration of Ca is high, because this is logically inappropriate.

After iteration through molecules involved in one or two reactions, the partial pathways (see Fig. 1) are

1.1, 1.2	1.3, 1.1	1.4, 1.5	2.2, 2.1	2.3
2.4	3.2, 3.1	4.2, 4.1	5.1, 5.2, 5.1	6.1, 6.2
7.1	7.2	8.1	8.2	8.3

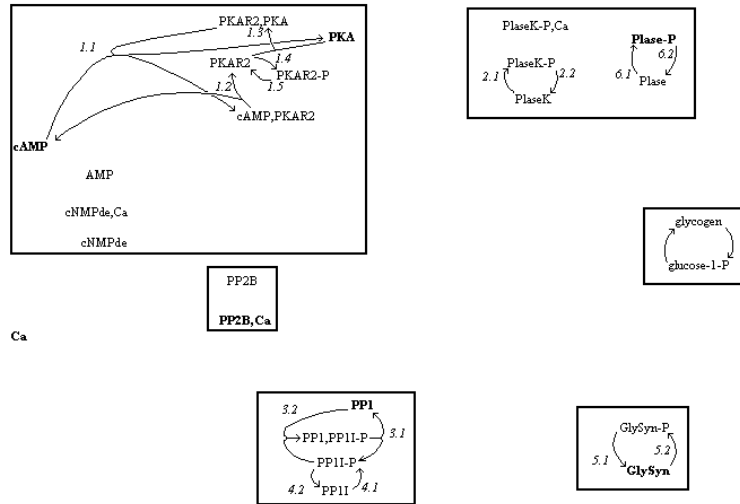


Fig. 1: Partial pathways involving two or more reactions, found after iteration over molecules involved in one or two reactions. All of these partial pathways are localized within functional blocks (rectangles).

After iterating for molecules involved in three or fewer reactions, the list of partial pathways is

1.1, 1.2	1.3, 1.1	1.4, 1.5	2.2, 2.1	2.3, 2.4
3.2, 3.1	4.2, 4.1	5.1, 5.2	6.1, 6.2	
c2.3, 6.2, 6.1	7.1, 7.2	8.1, 8.2	c8.1, 1.2, 8.3	

After iterating for molecules involved in four or fewer reactions, the list of partial pathways (see Fig. 2) is

1.1, 1.2, 1.3	1.1, 1.2, 1.4	1.3, 1.1	1.1, 1.2, (c7.1, 1.4, 1.5)
1.5, 1.3	1.5, 1.4	2.2, 2.1	2.3, 2.4
4.1, 3.2	c7.1, 4.1, 4.2	c7.1, 3.1, 4.2, 4.1	3.2, 3.1
c2.3, 6.2, 6.1	7.2	8.1, 8.2	5.1, 5.2
			c8.1, 1.2, 8.3

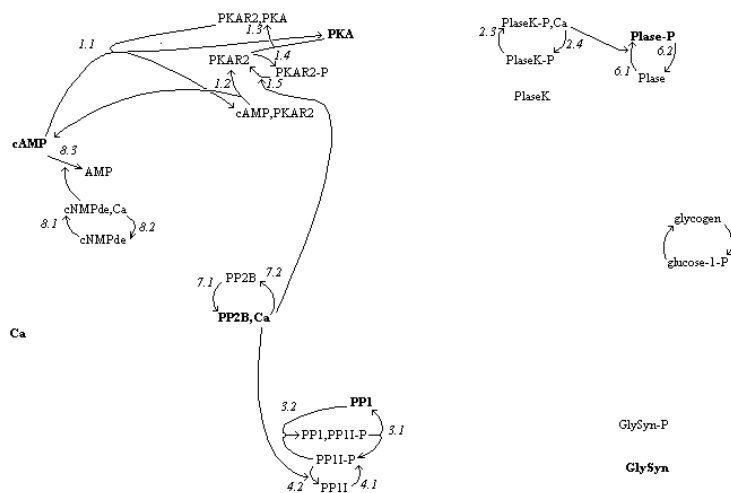


Fig. 2: Partial pathways involving two or more reactions found during iteration over molecules involved in three or four reactions. Most of these partial pathways are within functional blocks, but regulation of two blocks by PP2B,Ca is found.

After iterating for molecules involved in six or fewer reactions the partial pathways, shown in Fig. 3, are

- | | | | |
|----------------|----------------|---------------------|----------------------------|
| 1.1, 1.2, 1.3 | 1.1, 1.2, 1.4 | 1.3, 1.1 | 1.1, 1.2, (c7.1, 1.4, 1.5) |
| 1.5, 1.3 | 1.5, 1.4 | c1.1, 2.2, 2.1 | c3.1, 2.1, 2.2 |
| c3.1, 2.4, 2.2 | 2.1, 2.3 | 2.3, 2.4, 2.3 | 3.1, 3.2, 3.1 |
| c7.1, 4.1, 4.2 | c7.1, 3.1, 4.2 | c1.1, 4.2, 4.1, 3.2 | |
| c1.1, 5.1, 5.2 | c3.1, 5.2, 5.1 | c3.1, 6.1, 6.2 | |
| c2.1, 6.2, 6.1 | c2.3, 6.2, 6.1 | 7.1, 7.2 | 8.1, 8.2 |
| c8.1, 1.2, 8.3 | | | |

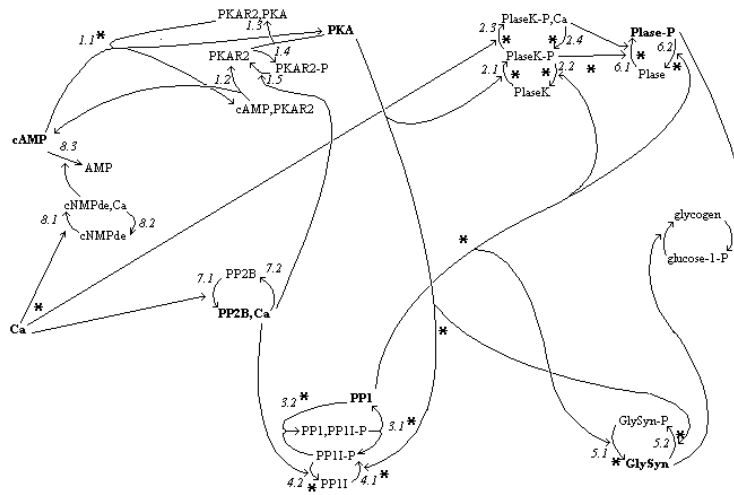


Fig. 3: All partial pathways. Those involving two or more reactions, found during iteration over molecules involved in five or six reactions, are marked *. Pathways of regulation by PKA and PP1 are found.

4.3 Use the partial pathways to construct responses to elevation of cAMP with low calcium, and to elevation of calcium with low cAMP

Knowing the partial pathways, one can sketch the consequences of changes in the concentration of cAMP or of Ca. The arguments depend on mass action: If the concentration of a supplied reactant is low, the reaction will proceed at a low rate, as will other reactions dependent on it. To find these reactions one can examine the input (left) sides of reactions in the reaction list, looking for reactants, catalysts, or cofactors that are in low concentration because a supplied reactant is in low concentration.

- Responses to elevation of cAMP with low Ca: When the Ca concentration is low, reactions 2.3, 7.1, 8.1 and their dependent reactions 1.5, 4.2, and 8.3 can be neglected. The remaining partial pathways are

1.1, 1.2, 1.3	1.1, 1.2, 1.4	c1.1, 2.2, 2.1	c3.1, 2.1, 2.2
c3.1, 2.4, 2.2	3.1, 3.2, 3.1	c1.1, 5.1, 5.2	c3.1, 5.2, 5.1
c3.1, 6.1, 6.2	c2.1, 6.2, 6.1		

cAMP acts through reaction 1.1, producing the observed partial pathways:

c1.1, 2.2, 2.1, (c2.1, 6.2, 6.1)	activate phosphorylase
c1.1, 4.2, 4.1, (3.2, 3.1, 3.2)	inhibit PP1
c1.1, 5.1, 5.2	inhibit glycogen synthase

Reactions 1.3 and 1.4 compete for PKAR2, with 1.4 dominant because 1.5 is negligible. Reaction 4.1 promotes 3.2 and so disfavors 3.1, so PP1 is not available to catalyze reactions 2.2, 5.1 and 6.2 .

- Responses to elevation of Ca with low cAMP: When the cAMP concentration is low, reaction 1.1 and its dependent reactions 2.1, 4.1, and 5.2 can be neglected. The remaining partial pathways are

c7.1, 1.4, 1.5	1.5, 1.3	1.5, 1.4	c3.1, 2.1, 2.2
c3.1, 2.4, 2.2	2.1, 2.3	3.1, 3.2, 3.1	c7.1, 4.1, 4.2
c7.1, 3.1, 4.2	c1.1, 4.2, 4.1, 3.2		c1.1, 5.1, 5.2
c3.1, 5.2, 5.1	c3.1, 6.1, 6.2	c2.1, 6.2, 6.1	c2.3, 6.2, 6.1
c8.1, 1.2, 8.3			

Ca acts through reactions 2.3, 7.1, and 8.1, producing the observed partial pathways.

c8.1, 1.2, 8.3	degrade cAMP
(c7.1, 1.4, 1.5), 1.3	release PKAR2, inhibit PKA
1.5, 1.4, 1.5	potentiate PKA activity
3.1, 3.2, 3.1	inhibit PP1
c7.1, 3.1, 4.2	inhibit PP1I
c7.1, 4.1, 4.2	inhibit PP1I
c2.3, 6.2, 6.1	activate phosphorylase
c3.1, 6.1, 6.2	inactivate phosphorylase
c3.1, 5.2, 5.1	activate glycogen synthase
c3.1, 2.1, 2.2	inhibit phosphorylase kinase
c3.1, 2.4, 2.2	inhibit phosphorylase kinase

4 Discussion

As the preceding example illustrates, the algorithm that Mavrovouniotis et al. (1990) devised to assemble metabolic pathways can construct pathways for signal transduction. In this example the reactions assembled in a unique pathway. However, the algorithm can find multiple pathways if they exist (*e.g. loc. cit.*, Appendix B). Mavrovouniotis et al. (1990) implemented their algorithm as a LISP program. The version discussed here is being implemented as a browser to visualize pathways (A. Wozniewicz, personal communication).

The preceding example implies that the algorithm can also assemble other intracellular networks, including pathways mediating gene regulation and cell motility, and multicellular networks, as in the development of an embryo and the operation of endocrine and immune systems. To use the algorithm in these contexts, one keeps track of all precursors for a process, whether they are modified in the process, catalyze it, regulate it, or serve as cofactors. The construction of partial pathways helps one to recognize functional clusters of processes and to appreciate the dynamics of their operation, at first intuitively and then through simulation.

In practice the algorithm is used differently for metabolism than for networks that modify enzymes. Metabolism transmits and transforms matter; the signaling net transmits and processes information. For metabolism the algorithm can provide correct stoichiometry of all metabolites. One can enforce stoichiometry to eliminate intermediate compounds, so that the starting molecules are wholly converted to the final products. In metabolism the forward and reverse reactions of a transformation are not used together, since this is a trivial way to make and remove intermediates. The extent to which the algorithm can provide stoichiometry if enzymes react is unclear. In signaling pathways the molecular inputs elicit changes in the concentrations of the output molecules but can leave intermediates altered in state. The forward and reverse reactions of a transformation may both function in transmitting information through a pathway.

As used here, the algorithm assembled pathways from specified real molecules and reactions. However, it is likely to be most useful in assembling possible pathways from databases of molecules and reactions, including hypothetical candidates. The molecules that participate in a reaction may be specified explicitly by listing them, as was done here. However, a more flexible and powerful way to characterize possible interactions among molecules is implicit specification, in terms of their binding sites. If a molecule has a specified kind of site it can bind to

any other molecule in which a complementary site is available, unless steric hindrance or another constraint prevents this. If the algorithm can connect molecules with complementary binding sites, it can assemble a greater variety of less obvious, potentially interesting pathways. At present the data on the identity, specificity, and complementarity of binding sites seem too sparse to use implicit specification, but this situation is changing rapidly.

References

1. J. E. Bailey, "Toward a science of metabolic engineering" *Science* 252, 1668-1675 (1991)
2. J. Happel, P. H. Sellers and M. Otarod, "Mechanistic study of chemical reaction systems" *Ind. Engng. Chem. Res.* 29, 1057 (1990)
3. D. G. Hardie, *Biochemical Messengers*. (Chapman and Hall, New York, 1991).
4. M. L. Mavrovouniotis, G. Stephanopoulos, and G. Stephanopoulos, "Computer-aided synthesis of biochemical pathways" *Biotechnol. Bioengng.* 36, 1119 (1990)
5. M. L. Mavrovouniotis, G. Stephanopoulos, and G. Stephanopoulos, "Synthesis of biochemical production routes" *Computers chem. Engng* 16, 605 (1992).
6. E. Melendez-Hevia, T. G. Waddell, and F. Montero, "Optimization of metabolism: The evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle" *J. theor. Biol.* 166, 201 (1994)
7. J. E. Mittenthal, B. Clarke, and M. Levinthal, in *Thinking About Biology, "Designing Bacteria"* Ed. W. D. Stein and F. J. Varela (Addison-Wesley, Reading, MA, 1993)
8. J. E. Mittenthal, A. Yuan, B. Clarke, A. Scheeline "Designing metabolism: Alternative connectivities for the pentose phosphate pathway" in preparation, 1996.
9. A. Seriossis and J. E. Bailey "MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways" *Biotechnol. Bioengng* 31, 587 (1988)