It is the lack of data to support a genetic or biochemical basis for BiDil's mechanism of action in African Americans that remains problematic. And it is this aspect of the work that we regard as scientifically dubious.

According to NitroMed's chief medical officer, Manuel Worcel, the company is currently analyzing data from ten genetic markers linked to heart failure, but cautions "we cannot rule out socioeconomic factors yet"[1]. Thus, on the basis of the evidence presented so far, BiDil has been approved for a group of people in a certain age group with a certain skin color whose predisposition to responding to the drug could be influenced by any number of vague social, economic, geographical, lifestyle and dietary factors. Access to care, treatment intensity, compliance and a host of other factors could just as easily be to blame.

Thus, although we welcome BiDil as an adjunct to other congestive heart disease therapies, such as angiotensin-converting enzyme (ACE) inhibitors and diuretics, we do not accept that BiDil is a model for targeted medicine. The way forward must be to find genetic or biochemical markers that are associated with response to therapy; to be sure, in certain cases, these markers will be abundant in certain racial groups. For example, deficiency in dihydropyrimidine dehydrogenase (DPD) is known to correlate with poor metabolism of fluorouracil, a standard chemotherapeutic treatment. The incidence of DPD in blacks is 9.4% compared with 0.9% in whites[1]. Such biochemical evidence provides a good rationale for race-based therapy.

In contrast, NitroMed's approach is to let the medicine go where the data take it. It is all very well to go on a fishing expedition by mining clinical data for patient populations (any population will do no matter how ambiguously defined) that correlate with drug response. But how useful is this in the absence of a molecular rationale or more certainty that the association is not spurious?

The question is: once you go down this road, where do you stop? Approvals for drugs in blondes with congestive heart failure perhaps? (Not BiDil, but BiMbo for her or HiMbo for him, peroxide blondes notwithstanding.) Or people of Welsh ancestry? Or people with a predilection for Welsh rarebit?

Already, in recent weeks, Seattle biotech company Cell Therapeutics announced it aims to get Xyotax, an encapsulated form of the chemotherapeutic paclitaxel, approved for treating women with advanced non-small cell lung cancer. The drug does not show significant benefit overall for patients with non-small cell lung cancer. Ditto in women with the cancer in two separate trials. It is only when the women's results from the two separate trials are combined that, hey presto, significance appears! According to a release from the company, it now wants to use data derived from trials where the drug was used in combination with others to justify a trial when the drug is used alone. The rationale for this sex-based analysis is apparently that estrogen present in women stimulates "an enzyme that breaks down the polymer," enabling the paclitaxel to kill the tumor.

This is not targeted medicine. It is data mining *in reductio ad absurdum*. With Herceptin (trastuzumab) and Erbitux (cetuximab), at least we know what the drug target is and exclusion or inclusion is based on the raised presence of that target. With Xyotax, no such target identified—just a post-hoc, nonrational, last-gasp analysis to save a drug that otherwise would be dead in the water.

1. Branca, M.A. *Nat. Rev. Drug Discov.* **4**, 615–616 (2005).

'On-the-fly' or 'generate-first' modeling?

To the editor:

In a paper in the January issue (*Nat. Biotechnol.* **23**, 131–136, 2005), Lok and Brent report methods and software, called Moleculizer, for automatically generating computational models of biochemical systems. Such rule-based modeling tools are needed to study systems marked by combinatorial complexity[1,2]. A feature of Moleculizer, which distinguishes it from related tools, such as BioNetGen[3,4] and BIOCHAM[5] (http://contraintes.inria.fr), is its implementation of 'on-the-fly' reaction network generation during Monte Carlo simulation of discrete-event reaction dynamics[6]. For reasons noted in the paper and elsewhere[1], this feature is a desirable capability, but we feel it necessary to mention drawbacks of the approach and advantages of an alternative approach, which did not receive adequate attention in the paper. Also, to abate confusion, we feel it necessary to point out that the suggestion of Lok and Brent to modify rate constants of reactions based on the molecular weights of reactants is founded on faulty premises.

The on-the-fly method closely ties network generation to simulation of network dynamics, which are governed by parameters, such as rate constants and initial population densities. Thus, the parameter values used in a simulation may determine the output of network generation. The parameter values can affect which elements of a network are important (e.g., which protein complexes are populated), and different network elements can be important under different conditions[7]. This situation presents a conundrum because parameter values are mostly unknown even for intensely studied systems.

Furthermore, the dependence of network generation on parameter specification and simulation can have undesirable consequences in practice. A Monte Carlo approach to simulation, which is an essential feature of Moleculizer, can be computationally expensive, as noted by Lok and Brent. Thus, use of Moleculizer in a procedure, for example, that involves repetitive simulation and variation of parameters, as in a parameter identification or sensitivity analysis routine, may be slow, because in general, network generation and concomitant Monte Carlo simulation must be performed each time parameters are varied. Reliance on inefficient computational procedures can be a barrier to comparing a model's predictions with experimental observations.

An alternative approach to on-the-fly network generation involves the use of reaction rules (referred to as reaction generators in the paper), which can be more general than the modules of Moleculizer[3–5], to produce a list of reactions without performing a simulation of network dynamics[1,3,4]. This approach was downplayed by Lok and Brent because of the potential for the size of a network to exceed the limits of computer memory. Despite this concern, which is arguably pessimistic, the 'generate-first' method has the advantage of being independent of both parameter specification and simulation.

Once a biochemical reaction network has been generated (not necessarily in its entirety), it can be used to formulate different types of models[3,4]. One can generate a system of coupled ordinary differential equations, a stochastic simulation algorithm (SSA) etc. The various rule-based modeling tools available can each be used to obtain these different types

of models, but use of Moleculizer unavoidably involves SSA-based Monte Carlo simulation.

Notably, the method of first generating a network and then simulating it has proven useful. For example, it has been used in model-based studies of early events in immunoreceptor signaling[2,8,9]. This approach has also been used to generate a model that is closely related to models discussed by Lok and Brent for yeast α-factor signaling. This model, available at our web site (http://cellsignaling.lanl.gov), was first mentioned in ref. 3 (as an example of BioNetGen capabilities) and is based on a scheme illustrated in ref. 1. It demonstrates that on-the-fly network generation is not always necessary. More work is needed to better understand the advantages and disadvantages of the two approaches, which we think are complementary.

Lok and Brent suggest a formula (on p. 135) for assigning rate constants to reactions. This formula is applied inappropriately, as we will discuss below, but its introduction is meant to account for the diffusivities of reactants, which depend on their molecular weights. Modification of rate constants based on molecular weights, which is an optional feature of Moleculizer, is an example of a context-sensitive model refinement, one that predicts how reactions of the same essential type are affected by varying molecular context. Here, context is variable because the molecular weights (and diffusivities) of reactants depend on association of binding partners. However, model modifications for this type of contextual variability, even with the use of applicable formulas, is unjustified or unnecessary in many cases.

Rates of reactions depend on the molecular weights (or equivalently, diffusivities) of reactants only when reactions are diffusion-limited; no corrections are needed or justified in reaction-limited cases. Furthermore, when diffusion is limiting (that is, much slower than chemical transformation), modifications of rate constants are expected to be minor in many typical situations[4]. For example, binding of a cytosolic protein to a membrane protein cannot be expected to significantly affect the diffusivity of the complex, because the viscosity of the cell membrane is far greater than that of the cytosol.

Finally, as mentioned earlier, the formula given by Lok and Brent is inapplicable for the types of reactions under consideration. The equation from which it is derived depends on the assumption of an ideal gas[10] (also see ref. 7 of the paper). In fact, the underlying basis for the formula is the kinetic theory for an ideal gas. Applicable formulas can be derived from diffusion theory and used if refinements of the kind suggested by Lok and Brent are needed[4]. BioNetGen now implements two methods of on-the-fly network generation, the method described by Lok and Brent and a closely related method described in ref. 4.

Michael L. Blinov, James R. Faeder, Jin Yang, Byron Goldstein & William S. Hlavacek

Theoretical Biology and Biophysics Group, Theoretical Division, Mail Stop K710, Group T-10, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.
e-mail: bionetgen@lanl.gov

1. Hlavacek, W.S., Faeder, J.R., Blinov, M.L., Perelson, A.S. & Goldstein, B. *Biotechnol. Bioeng.* **84**, 783–794 (2003).
2. Goldstein, B., Faeder, J.R. & Hlavacek, W.S. *Nat. Rev. Immunol.* **4**, 445–456 (2004).
3. Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. *Bioinformatics* **20**, 3289–3291 (2004).
4. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Complexity*, **70** , 22–41 (2005).
5. Fages, F., Soliman, S. & Chabrier-Rivier, N. *J. Biol. Phys. Chem.* **4**, 64–73 (2004).
6. Faulon, J.-L. & Sault, A.G. *J. Chem. Inf. Comput. Sci.* **41**, 894–908 (2001).
7. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Syst. Biol.,* **2**, 5–15 (2005).
8. Goldstein, B. *et al. Mol. Immunol.* **38**, 1213–1219 (2002).
9. Faeder, J.R. *et al. J. Immunol.* **170**, 3769–3781 (2003).
10. Atkins, P.W. in *Physical Chemistry*, 3rd edn., 655–657 (W.H. Freeman and Company, New York, 1986).

Lok and Brent respond:

Our simulation program, Moleculizer, calculates cellular reaction networks by allowing protein complexes to form *in silico*. From the correspondence of Hlavacek and colleagues, we take two main points. They argue that there are advantages to generating reaction networks beforehand, followed by running the simulation to solve them. They also point out that the formulation we use to estimate new intracellular reaction rates is likely to be an oversimplification.

We agree with these points, and we discussed them both extensively in our published paper. Using Moleculizer or other rule-based programs to generate reaction networks that are then fed into the other kinds of simulators has, in some cases, the advantage of decreasing computational cost. We envisioned that the use of Moleculizer to generate networks solved by other simulators would be perhaps its best use in the future. We discussed this explicitly in the article; among other things, it is why we placed so much work and emphasis on enabling export of reaction networks via systems biology markup language (SBML). Similarly, we recognized that our formula for calculating intracellular diffusion is at best a simplification—in the paper, we refer to it as a "placeholder"—and stressed in the discussion that the modular nature of the existing code makes it easy for users to experiment with other, perhaps more complex, formulae that might give better results. To the extent that Hlavacek and colleagues suggest we did not consider or discuss these points in our published work, we believe that these authors are attacking a straw man, perhaps to focus attention on their own forthcoming work on simulation of cellular reaction networks.

We wonder if another trigger for their correspondence may have been a difference in scientific cultures. Moleculizer and other ongoing simulation work arise from a biological research effort, the Alpha project, whose ultimate goal is to predict the behavior of a single, extremely well-characterized, signal transduction pathway in yeast. Just like the biological experimentation to which it is coupled, Moleculizer is a work in progress; at the end of the day, we view it as a tool. This view enables us to write modular code to solve problems simply and replace those solutions with more sophisticated ones as the need arises. When we wish to compute a reaction network, we are comfortable beginning with a computationally inefficient route until it becomes too burdensome to follow further. When we wish to accommodate molecular diffusion, we are comfortable beginning with a simple formula until such time as the results from using it diverge unacceptably from those obtained by measurements of the living system. In contrast, Hlavacek and colleagues are physicists in a theoretical department. In some areas of physics, 'theory' is relatively more important, and there may be a tendency to try to get the theoretical basis right from first principles, rather than being resigned to introducing, modifying and discarding ideas and formulas as one goes along. Neither stance is more 'correct' than the other. However, we believe that, for a good deal of the work that needs to be done to compute the behavior of biological systems, concentration on building a 'perfect' simulator may not be as important as production of supple (and complex!) code that can handle the significant challenges, including the myriads of different protein complexes, posed by living systems, and that can be continually modified as tight coupling to ongoing experimentation produces new challenges and results. For at least some applications of simulation to biology, the perfect may be the enemy of the good.

Larry Lok & Roger Brent

The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704, USA.
e-mail: lok@molsci.org