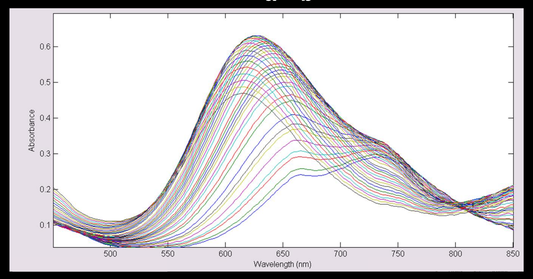
Defining the Relationship: Computer-Driven Characterization of the Binding of Host and Guest Molecules

Joyce A. Chew, Nathanael P. Kazmierczak, Dr. Douglas A. Vander Griend. Calvin College, Grand Rapids, Michigan

Introduction

$$aH + bG \rightleftharpoons H_aG_b \quad \Delta G^\circ = x$$



Three questions:

- How many species form?
- How strongly do they bind?
- What is the composition of those species?

We can determine the number of species through principal component analysis, and given a chemical model (the composition of those species), we can calculate ΔG°, but model searching is traditionally done by hand.

Methods

Framing the Problem

Instead of optimizing absolute stoichiometries, we reparameterize to ratios and oligomerities.

- For H_4G_6: ratio = 1.5, oligomerity = 2

For a system with n reactions, we actually optimize 2n parameters (ratios / ΔG°, oligomerities / ΔG°). We eliminate permutation ambiguity when optimizing ratios by "chaining" reactions or by enforcing nonlinear constraints.

Optimization approaches

We adopt two viewpoints to select optimization algorithms: "numerical" and "chemical".

- Numerical ("continuous" ratios): Levenberg-Marquardt (gradient descent), PSWARM (particle swarm/pattern search hybrid)
- Chemical ("discrete" ratios/oligomerities): MADS (mixed-integer adaptive mesh)

Algorithm Comparison

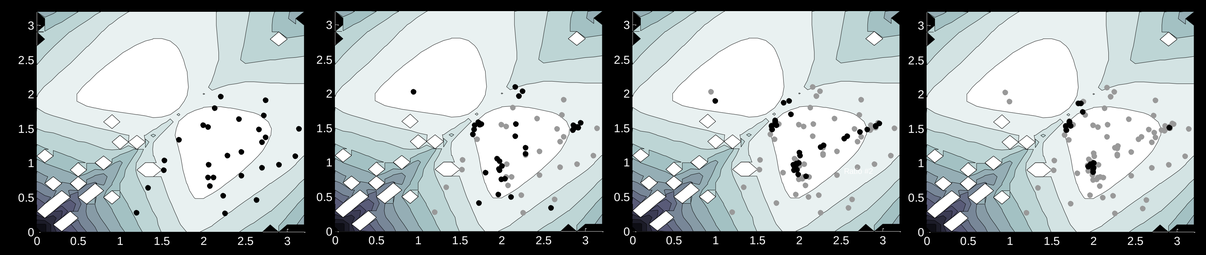


Figure 1: *Points tested by multi-start Levenberg-Marquardt: multi-start gradient descent over continuous values*

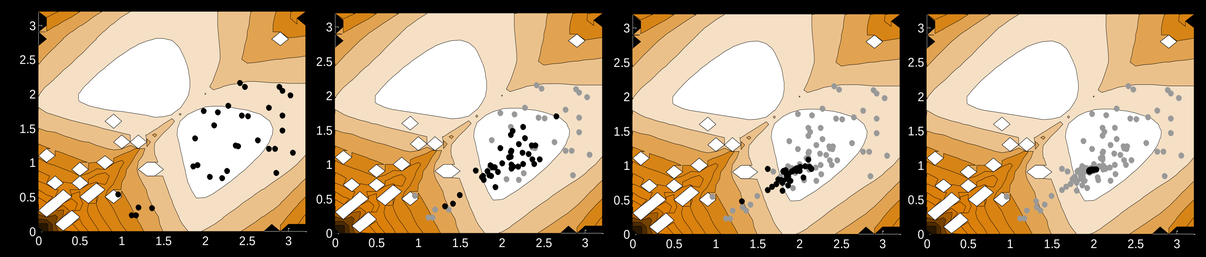


Figure 2: *Points tested by PSWARM: global optimization governed by stochastic swarm behavior (inertia, social, cognition) over continuous values*

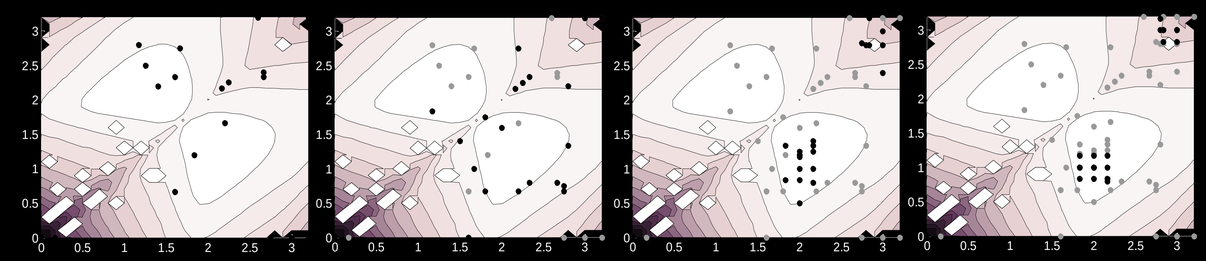


Figure 3: *Points tested by one instance of MADS: mixed-integer semi-global optimization*

System	Algorithm	Converged points	Time	Optimized ratios	Final RMSR
$2H + G \rightleftharpoons H_2G$	Levenberg-Marquardt	62/100	137s	0.49994	8.1601e-06
	PSWARM	-	57s	0.49994	8.1601e-06
	MADS	1/10	26s	0.5	6.1255e-08
$2H + G \rightleftharpoons H_2G$ $H + G \rightleftharpoons HG$ $HG + G \rightleftharpoons HG_2$	Levenberg-Marquardt	36/100	359s	0.5067, 1.0875, 1.9986	5.0139e-04
	PSWARM	-	287s	0.5025, 1.2271, 2.0063	3.0843e-04
	MADS	1/30	466s	0.5, 1, 2	7.4313e-04

Conclusions

Algorithm comparison

MADS produces intuitively chemically meaningful results, but it is slow and not always accurate, especially in larger systems.

- True ratios: 0.5, 1, 2, 3, 4
  - MADS (20 min): 0.6, 1, 1.33, 2.8, 3.67
  - PSWARM (9 min): 0.51, 1.11, 1.94, 2.94, 3.96

MADS appears to prioritize optimizing ΔG° rather than changing stoichiometries, which change the meaning of the chemical model much more. Therefore, the best approach is to run an initial PSWARM, then generate the closest chemically meaningful models.

Reformulation of the problem

The assumption of one host molecule (in "numerical" viewpoint) improves convergence, but biases the ratios that are less than one.

References

Levenberg, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Q. Appl. Math. 1944, 2 (2), 164–168.

Vaz, A. I. F.; Vicente, L. N. A Particle Swarm Pattern Search Method for Bound Constrained Global Optimization. J. Glob. Optim. 2007, 39 (2), 197–219.

Le Digabel, S. Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm. ACM Trans Math Softw 2011, 37 (4), 44:1–44:15.

Acknowledgements

This work was supported by the National Science Foundation (NSF) through an REU grant from the Division of Chemistry (1310402) and through an S-STEM grant. We are additionally grateful to the Arnold and Mabel Beckman Foundation for supporting Nathanael Kazmierczak as a Beckman Scholar, and to the Jack and Lois Kuipers Applied Mathematics Endowment for supporting Joyce Chew.