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

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

Introduction

\[ aH + bG \neq H_aG_b \quad \Delta G^\circ = x \]

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 \( \Delta G^\circ \), but model searching is traditionally done by hand.

Methods

Framing the Problem

Instead of optimizing absolute stoichiometries, we reparameterize to ratios and oligomerities.
- For \( H_aG_b \) ratio = 1.5, oligomerity = 2

For a system with \( n \) reactions, we actually optimize \( 2n \) parameters (ratios / \( \Delta G^\circ \), oligomerities / \( \Delta G^\circ \)). We eliminate permutation ambiguities 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

<table>
<thead>
<tr>
<th>System</th>
<th>Algorithm</th>
<th>Converged points</th>
<th>Time</th>
<th>Optimized ratios</th>
<th>Final RMSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2H + G \neq H_2G )</td>
<td>Levenberg-Marquardt</td>
<td>62/100</td>
<td>137s</td>
<td>0.49994</td>
<td>8.1601e-06</td>
</tr>
<tr>
<td></td>
<td>PSWARM</td>
<td>-</td>
<td>57s</td>
<td>0.49994</td>
<td>8.1601e-06</td>
</tr>
<tr>
<td></td>
<td>MADS</td>
<td>1/10</td>
<td>26s</td>
<td>0.5</td>
<td>6.1255e-08</td>
</tr>
<tr>
<td>( 2H + G \neq H_2G )</td>
<td>Levenberg-Marquardt</td>
<td>36/100</td>
<td>359s</td>
<td>0.5067, 1.0875, 1.9896</td>
<td>5.0139e-04</td>
</tr>
<tr>
<td></td>
<td>PSWARM</td>
<td>-</td>
<td>287s</td>
<td>0.5025, 1.2271, 2.0063</td>
<td>3.0843e-04</td>
</tr>
<tr>
<td></td>
<td>MADS</td>
<td>1/30</td>
<td>466s</td>
<td>0.5, 1, 2</td>
<td>7.4313e-04</td>
</tr>
</tbody>
</table>

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 \( \Delta G^\circ \) rather than changing stoichiometries, which changes 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


Acknowledgements

This work was supported by the National Science Foundation (NSF) through an REU grant from the Division of Chemistry (1100929) 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.

Framed by the Problem, we can determine the number of species. We adopt two viewpoints to select optimization algorithms: "numerical" and "chemical".

Optimization approaches

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

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