1. Supplementary Materials

The following sections contain the information and data deemed supplementary to the main text.

1A. Force Field Selection. The selection of the molecular force field is one of the most important pieces of any molecular simulation, and as such, it is critical to determine that the force field is able to reproduce certain key aspects of the physics in question. Cheong and Boon investigated the validity of using Charmm27, general AMBER, OPLS-AA/L, and Gromacs53a6 force fields for α-glycine crystal growth simulations[1]. The use of five different charge sets (CNDO, LCAU, DNP, DZP, 6-31 G*) was also investigated. Finally, the SPC, SPC/E, and TIP3P water models were investigated. Cheong and Boon conclude that the general Amber force field (GAFF), coupled with CNDO charges, is the optimal force field for the study of α-glycine crystal growth, as it was able to reproduce solution properties and the heat of solution well. Of the force field and charge sets questioned, the GAFF/CNDO combination was the only force field/charge combination that predicted a positive enthalpy of solution, meaning if any of the other force field/charge combinations were employed, crystallization would be an endothermic process, and would not take place. This is reflected in the work of Banerjee and Briesen who performed MD simulation of the (110) face of α-glycine using the Gromos53a6 force field in SPC water[2]. Even at very high supersaturations, crystal dissolution was observed. This reflects the degree of peril molecular modelers undertake when performing nucleation simulations for the purpose of determining either crystal growth kinetics, or nucleation kinetics and solubility, as there is a large degree of variation in the predicted solubility between force field sets. We elected to employ the GAFF force field with CNDO charges for glycine, the SPC/E water model for water, as this was the only force field/charge combination tested that yielded a positive enthalpy of solution, and the possibility of crystal growth. Thus, the simulations performed in this work were done with no knowledge apriori of the solubility of either polymorph predicted by this force field. Only the

knowledge that crystal growth would take place for α-glycine, reflected in the positive heat of solution for α-glycine predicted by Cheong and Boon was present.

1B. MD Computational Details. To minimize finite size effects, such as cluster-cluster interactions between periodic images, and fluid phase enrichment or depletion due to solid dissolution or growth, large scale MD was employed with systems of ~100K atoms and box lengths of ~10 nm. To create an initial configuration of molecules for fluid equilibration, 2700 glycine molecules were solvated with SPC/E water using the gromacs/5.0[3] solvate command at a concentration of 0.34 g/ml. The equilibrated fluid was then prepared by first performing MD equilibration in the NVT ensemble for 6 ns, and then equilibrating the fluid in the NPT ensemble for an additional 6 ns at 380 K. The final equilibrated concentration of the system 0.36 g /ml. A single α-, β-, or γ-glycine spherical nanonanocrystal, with varying radius, was embedded in the equilibrated fluid using the gromacs/5.0 solvate command. The energy of the system was then minimized in gromacs/5.0, using steepest descent minimization, with an energy minimization tolerance of 100.0 kJ/mol/nm.

All subsequent MD simulations were performed using an MD code consisting of LAMMPS[4] and NAMD[5] subroutines. Specifically, the integrator, thermostat, barostat, and real space force routines were taken from LAMMPS, and the PME solver was taken from NAMD. However, the parallelization was altered in that single MD trajectories are restricted to CPU sockets, using MPI, and are parallelized over the socket cores, and an Intel® Xeon Phi™ coprocessor, using OpenMP[6]. This allowed us to perform all MD trajectories simultaneously in an embarrassingly parallel manner. Furthermore, parallelizing over Intel® Xeon Phi™ coprocessors greatly reduced the number of CPU cores necessary to perform simulations. A full account of the MD optimizations

and parallelization schemes performed, along with a link to the full source code, can be found in the literature[6].

Independent MD trajectories were launched by generating randomly initialized velocities according to a Maxwell distribution. Short-range non-bonded forces were truncated at 1.2 nm. Reciprocal space electrostatic forces were computed using a particle mesh Ewald (PME) solver with a 1.0e-6-error tolerance. 100 grid points were used in each direction for the PME solver. The equations of motion were solved using a velocity Verlet algorithm with a 3.0 fs timestep. The temperature was controlled using a Langevin thermostat with a relaxation time of 2000 fs. The isotropic Berendsen barostat was used with a 1000 fs relaxation time and a set point of 1 atm. The shake algorithm was employed to constrain the motion of all bonds containing hydrogen. A shake tolerance of 1.0e-4 was employed for all simulations, and a maximum number of 25 shake iterations were allowed. Periodic boundary conditions were employed in all directions.

1C. Number of Particles in the Cluster. Solid and liquid particles were differentiated based off the local density of a molecule[7–9]. Glycine, a small amino acid, has a sufficiently high packing density (~1.61g/cm^3) to differentiate between liquid and solid molecules without reference to orientation ordering. In this work, molecules were considered solid if it possessed 11 or more neighbors in a 0.6 nm radius sphere[7]. That distance roughly corresponds to the first minimum in the glycine-glycine radial distribution function[7]. This number of neighbors corresponds to the tail end of the bulk glycine density distribution. Surface molecules were included in the clustering algorithm by counting a molecule as solid if it had a neighbor who met the density criterion, but did not meet the density criterion itself.

1D. Temperature Screening Computational Methodology. To perform temperature screenings, ten MD trajectories were launched from each of the initial polymorph specific nanocrystal seeds,

obtained from the Cambridge Structural Database[10] for up to 9 ns. All ten trajectories were then averaged to yield the average nucleus size drift data. Trajectories were launched at many temperatures until average drift was approximately zero. This required determining the melting temperature to within 1.25 K in some instances. Below the melting temperature estimate, clusters grew on average, and above the melting temperature, clusters dissolved on average. If different temperatures produced drift data that showed initial growth, followed by no growth or dissolution on average, the temperature that produced the least initial growth was selected as the melting temperature, as the critical size estimate obtained using this method is closest to the initial embedded cluster size. The y intercept, obtained from the linear fit to the averaged drift data at the critical temperature, was used to estimate the critical size. The interface equilibration was not used for the linear fit critical size analysis.

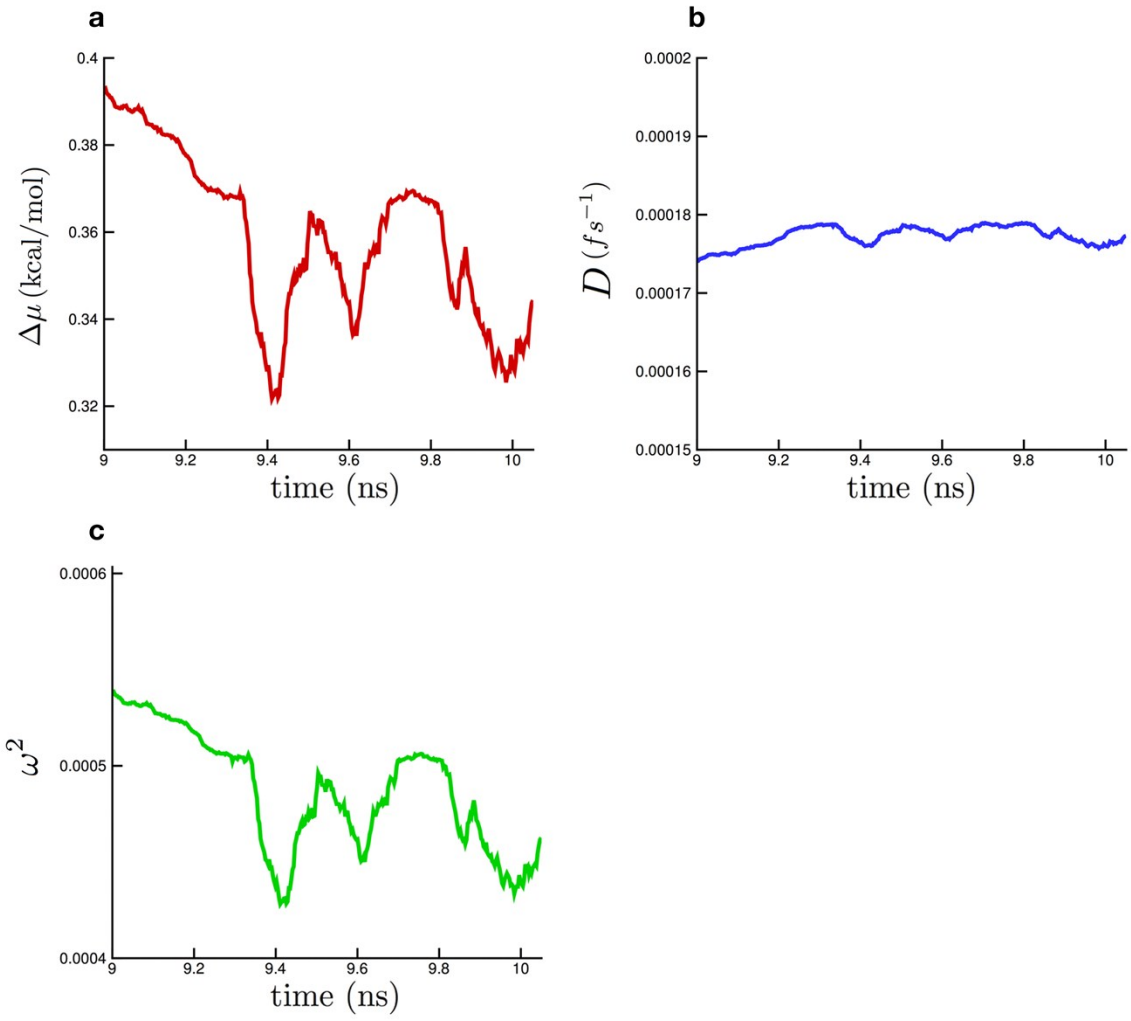
1E. MSD analysis and Parameter Estimation. For model MSD analysis, 10 clusters at the critical size were selected from the swarm of trajectories generated at the critical temperature. 20 trajectories were then launched from each cluster, for a total of 200 trajectories. Each trajectory was initiated by selecting momenta randomly from a Maxwell Boltzmann distribution. Trajectories were allowed to equilibrate for 750 ps after velocity initialization, and then MSD data was recorded.

In order to estimate the D, ω^2, and $\Delta\mu$ parameters from the MSD data, a least-squares error function was optimized using a Nelder-Meade downhill simplex method [11] implemented in MATLAB's fminsearch function. Because parameter estimates varied depending on the extent of the simulation, parameter estimates were tracked from 0 ns to the final simulation time, T_{final}, to determine when the parameters had converged with respect to simulation length for 1.05 ns. Parameters were averaged over this window. This resulted in parameters were estimated over

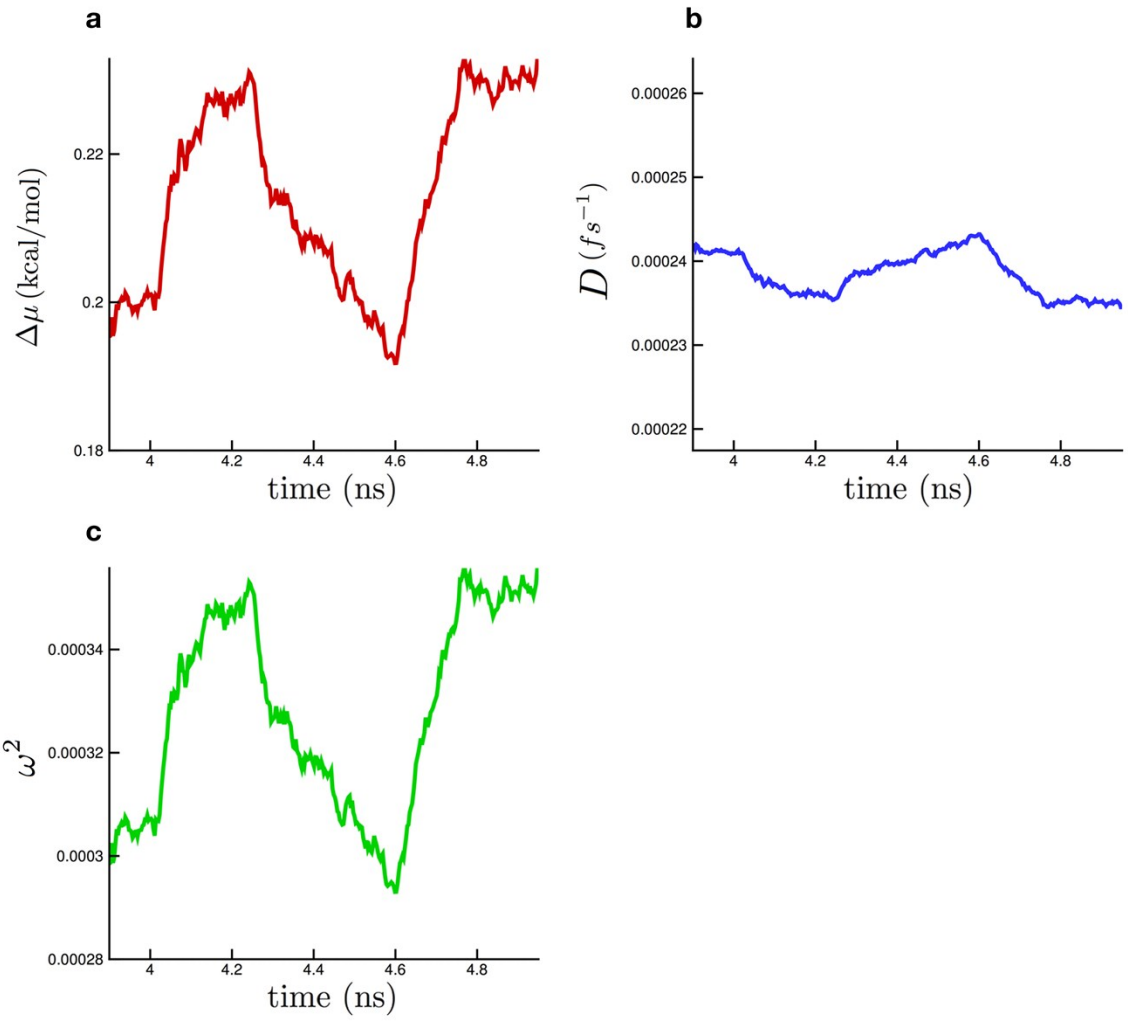
values of T_{final}=9.0 ns to T_{final}=10.05 ns for α-glycine, and T_{final}= 3.9 ns to T_{final}=4.95 ns for β-glycine. The converged parameters for α- are plotted in Supplementary figure 1, and the converged parameters for β- are plotted in Supplementary figure 2.

To perform Einstein MSD analysis, where the classical Einstein relationship is used to estimate the diffusivity of particles, 5 clusters at the critical size were selected from the swarm of trajectories at the critical temperature. 20 trajectories were then launched from each cluster for a total of 100 trajectories. Each trajectory was launched by selecting momenta randomly from a Maxwell Boltzmann distribution. Trajectories were allowed to equilibrate for 750 ps after velocity initialization, and then MSD data was recorded. Einstein MSD simulations were conducted for 600 ps. A linear fit to the data was then used to calculate the diffusivity, as required by equation (6). The linear fit data for Einstein diffusivity calculation is shown in Supplementary figure 3 and 4.

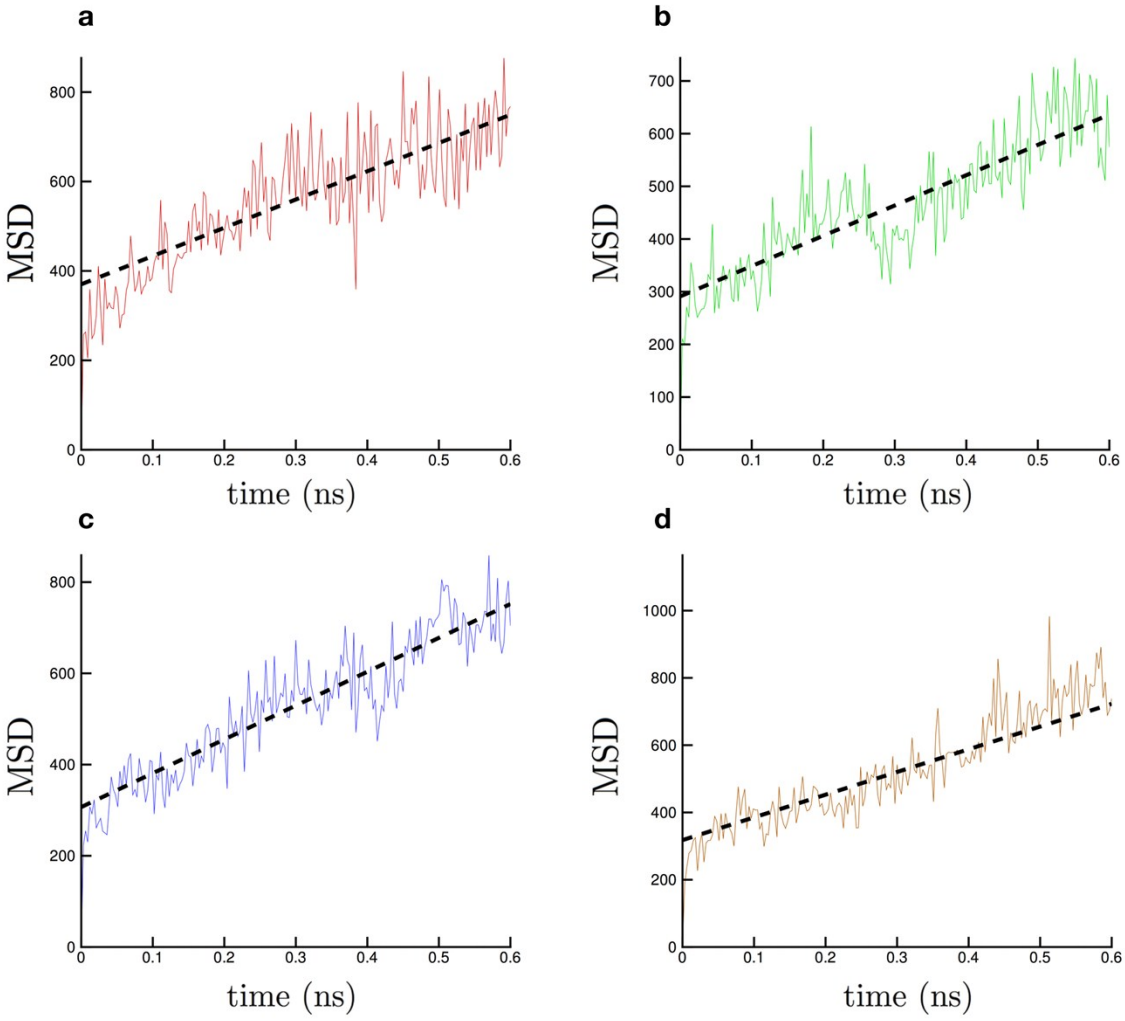
Supplementary Figure 1: The model MSD parameters are plotted again total simulation time (ns) for the case of α-glycine. In figure a, $\Delta\mu$ (kcal/mol) is plotted against time (ns). In figure b, D (fs^{-1}) is plotted against time (ns). In figure c, ω^2 is plotted against time (ns).



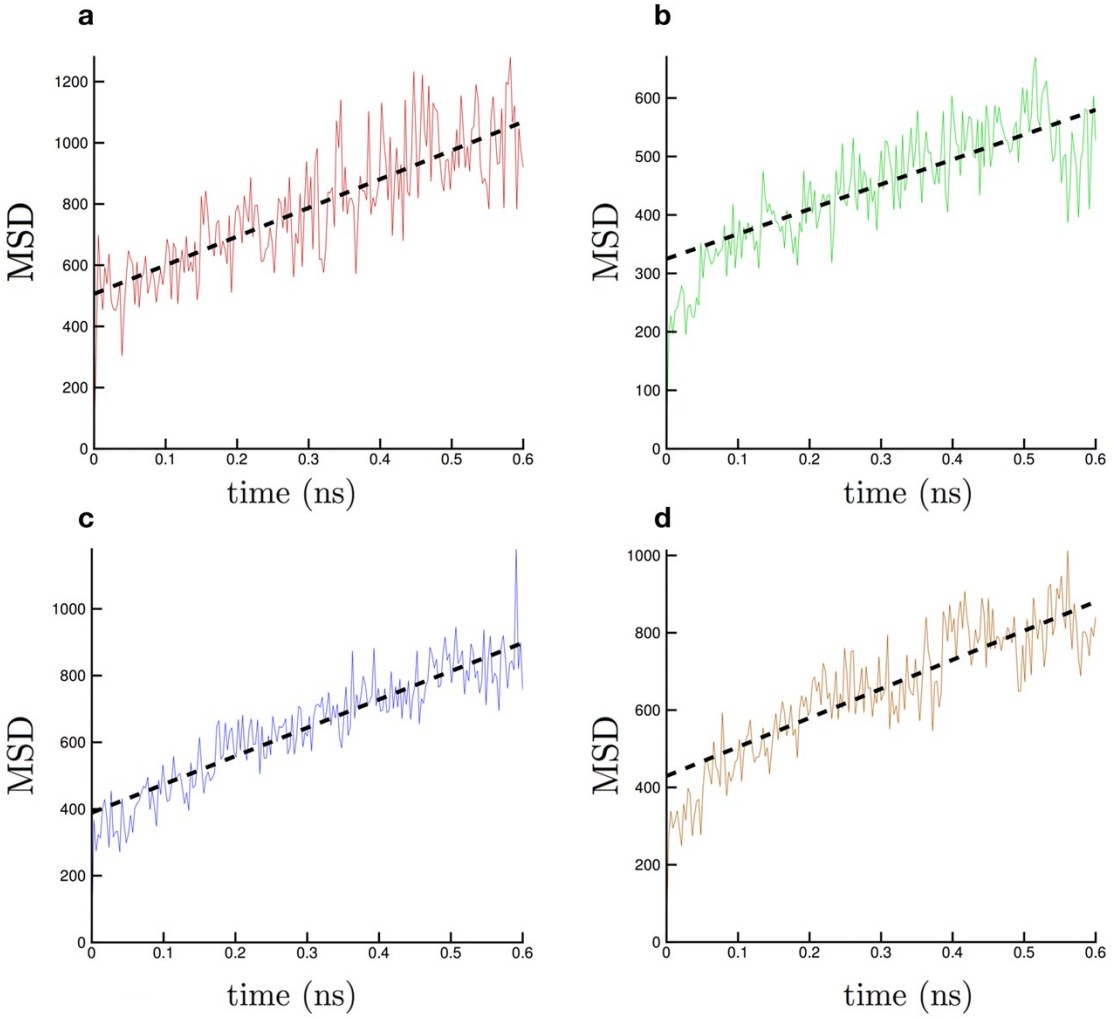
Supplementary Figure 2: The model MSD parameters are plotted again total simulation time (ns) for the case of β-glycine. In figure a, $\Delta\mu$ (kcal/mol) is plotted against time (ns). In figure b, D (fs^{-1}) is plotted against time (ns). In figure c, ω^2 is plotted against time (ns).



Supplementary figure 3. α-glycine MSD data for Einstein diffusivity coefficient determination for the initially 1.5nm particle (a), 2.0nm particle (b), 2.15nm particle (c), and 2.70nm particle (d), at their respective critical temperatures.



Supplementary figure 4. β-glycine MSD data for Einstein diffusivity coefficient determination for the initially 1.5nm particle (a), 2.0nm particle (b), 2.15nm particle (c), and 2.70nm particle (d), at their respective critical temperatures.



Supplementary Information References

1 D. W. Cheong and Y. Di Boon, *Cryst. Growth Des.*, 2010, **10**, 5146–5158.

2 S. Banerjee and H. Briesen, *J. Chem. Phys.*, 2009, **131**, 184705.

3 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

4 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

5 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

6 C. Parks, L. Huang, Y. Wang and D. Ramkrishna, *Mol. Simul.*, 2016.

7 N. Duff and B. Peters, *J. Chem. Phys.*, 2011, **135**, 134101.

8 G. Lanaro and G. N. Patey, *J. Phys. Chem. B*, 2015, **119**, 4275–4283.

9 C. Parks, A. Koswara, H. Tung, N. Nadkishor, B. Shailendra, Z. K. Nagy and D. Ramkrishna, *J. Mol. Pharmaeceutics*, 2016. (submitted)

10 F. H. Allen, *Acta Crystallogr. Sect. B Struct. Sci.*, 2002, **58**, 380–388.

11 J. C. Lagarias, J. A. Reeds, H. Wright and P. E. Wright, *J. Optim.*, 1998, **9**, 112–117.