[1].

[illegible]

[4]. But to

[illegible]

[6]0. [8] alternative selections. Here we show how genetic algorithms[5], mimicking biological evolution (survival [illegible]

[illegible]

[6] in solid-state physics[7], structural inorganic chemistry[8] and metallurgy[9]: how to

[illegible]

Even for a binary A_nB_m solid described by a unit cell with N sites, there can be $\sim 2^N$ possible structures (denoted as σ in the following). A direct quantum-mechanical evaluation of the total energy, $E_{QM}(\sigma) = \langle\Psi|H(\sigma)|\Psi\rangle$ (where Ψ is the wavefunction and $H(\sigma)$ is the electronic hamiltonian), of all 2^N configurations σ in

[illegible]

required for such a fit. If, on the other hand, we work with a smaller number is accepted on the basis of its fitness, representing the ability of this combination of MBITs to predict (through equation (1)) the energy of new configurations as determined from $E_{QM}(\sigma)$.

In Fig. 3a–c, we demonstrate the ability of the method to find the correct MBITs for three examples. We first tackle a case for which the underlying interactions $\{J\}$ are known exactly. We accomplish this by substituting for $\{E_{QM}(\sigma)\}$ a set of 60 energies $\{E_{CE}(\sigma)\}$ generated from a known set of MBITs[21] (Fig. 3). The energies $E(\sigma)$ are passed to the genetic algorithm to recover the five underlying MBITs from a pool of 45 possible candidates. Figure 3a shows the convergence to and correct identification of the exact MBITs for which both cross-validation score and fitting error are zero. We recover the right set of MBITs in 75 genetic algorithm steps (generations), after evaluating only 975 actual cluster expansions among 1.22 million possible combinations of MBITs.

In Fig. 3b, we turn to the application to actual first-principles data $\{E_{QM}(\sigma$