

Where are nature's missing structures?

GUS L. W. HART

Department of Physics & Astronomy, Brigham Young University, Provo, Utah 84602, USA
e-mail: gus.hart@gmail.com

Published online: 11 November 2007; doi:10.1038/nmat2057

Our society's environmental and economic progress depends on the development of high-performance materials such as lightweight alloys, high-energy-density battery materials, recyclable motor vehicle and building components, and energy-efficient lighting. Meeting these needs requires us to understand the central role of crystal structure in a material's properties. Despite more than 50 years of progress in first-principles calculations, it is still impossible in most materials to infer ground-state properties purely from a knowledge of their atomic components—a situation described as 'scandalous' in the well-known essay by Maddox¹. Many methods attempt to predict crystal structures and compound stability, but here I take a different tack—to infer the existence of structures on the basis of combinatorics and geometric simplicity². The method identifies 'least random' structures, for which the energy is an extremum (maximum or minimum). Although the key to the generic nature of the approach is energy minimization, the extrema are found in a chemistry-independent way.

A fundamental goal of solid-state physics and materials science is predicting the existence and properties of materials on the basis only of a knowledge of their constituents. For example, selecting two elements A and B from the periodic table, we should be able to predict whether or not they will form a stable compound or unmix. Furthermore, we should be able to predict the properties of the resulting compound or solid solution. To achieve this goal, we must be able to accurately treat the quantum-mechanical interactions of the constituent atoms in a solid. But in reality much of the progress towards achieving this goal has been made without a direct solution of the quantum-mechanical problem.

Early work based on the ideas of Hume-Rothery³, chemical arguments based on electronegativity^{4,5} and other empirical approaches were quite successful in the beginning of this endeavour. These approaches form the underpinnings of modern-day materials science. Conceptually similar but more modern versions of this approach are the Pettifor maps⁶, data mining⁷ and information theoretic⁸ methods. All of these approaches make predictions broadly consistent with known structures and compounds, and in many cases, they provide simple explanations for their existence. More importantly, they sometimes suggest unsuspected stable compounds.

There is another, less heuristic class of methods that attempts to make similar prognoses by working directly or indirectly from quantum mechanics. Included in this class are cluster-expansion methods^{9–11}, the coherent potential approximation methods and its many extensions^{12,13}, bond order potentials¹⁴ and the embedded-atom method^{15,16}.

Here I advocate another approach, an approach that suggests new structures simply by geometric simplicity. These structures have few atoms in the unit cell and simple atom–atom correlations.

For example, enumerating the possible binary structures of a face-centred-cubic (f.c.c.) lattice, we immediately find that the simplest structure is the well-known $L1_0$ structure. This structure is a very common intermetallic structure—exhibited by nearly 50 compounds. This simple structure can be described using a shorter set of basis vectors, and with fewer atoms per unit cell, than any other structure, and it can be completely specified by the atom–atom correlations of only the nearest and next-nearest neighbours.

In this approach, I assert that the geometric simplicity of a structure is an indicator of its likelihood. The approach proceeds as follows. First, I enumerate all possible structures (countably infinite) starting with those with the simplest descriptions. Then I look for their appearance in the experimental literature. Although the list of structures generated by this counting approach includes many well-known structures¹⁷, it also includes a number of 'missing' (as-yet-unobserved) structures—where are they? A likely candidate may have no experimental manifestation, not because it is not a solution to the configurational problem, but simply because it has been overlooked or not yet synthesized. Thus, this new structure may be an opportunity to fabricate new alloys or compounds.

Here, I illustrate the approach with a single example: binary compounds of an f.c.c. lattice. The results of two more examples indicate the generic nature of the approach. The combinatorial algorithm for counting is described in refs 18,19 and in the Methods section. Enumerating all structures of the type $A_{1-x}B_x$, we find only 17 with 4 or fewer atoms per unit cell, as shown in Fig. 1. Among the infinite number of possible f.c.c. structures, these are the simplest. Of these 17 structures, 9 are found experimentally, whereas 8 have not yet been observed—we are left to question why. Are such structures unlikely to exist generally? Or is it just the case that they have not yet been fabricated or recognized¹⁷?

The structures that have not been observed fall into two subclasses, those which have been predicted to exist and those for which no predictions have been made (yellow and purple shading, respectively, in Fig. 1). Armed with this information, we can now compare those that exist with those that are suspected to exist, and with those for which nothing is known, to make concrete arguments about why some of the 'candidate' structures may actually be physically realizable and why some may not be.

The key, the operating principle, it turns out is as follows: in essence, structures that have known prototypes strongly favour, at each bond length, just one type of bond, either unlike bonds (bonds between A and B atoms) or like bonds (A–A/B–B bonds). Depending on the nature of the A and B atom chosen for a particular binary system, chemistry will dictate that either like or unlike bonds are stronger (that is, energy lowering). Depending on the atoms involved, either kind of bonding can be energy lowering and thus either can lead to compound formation, but one

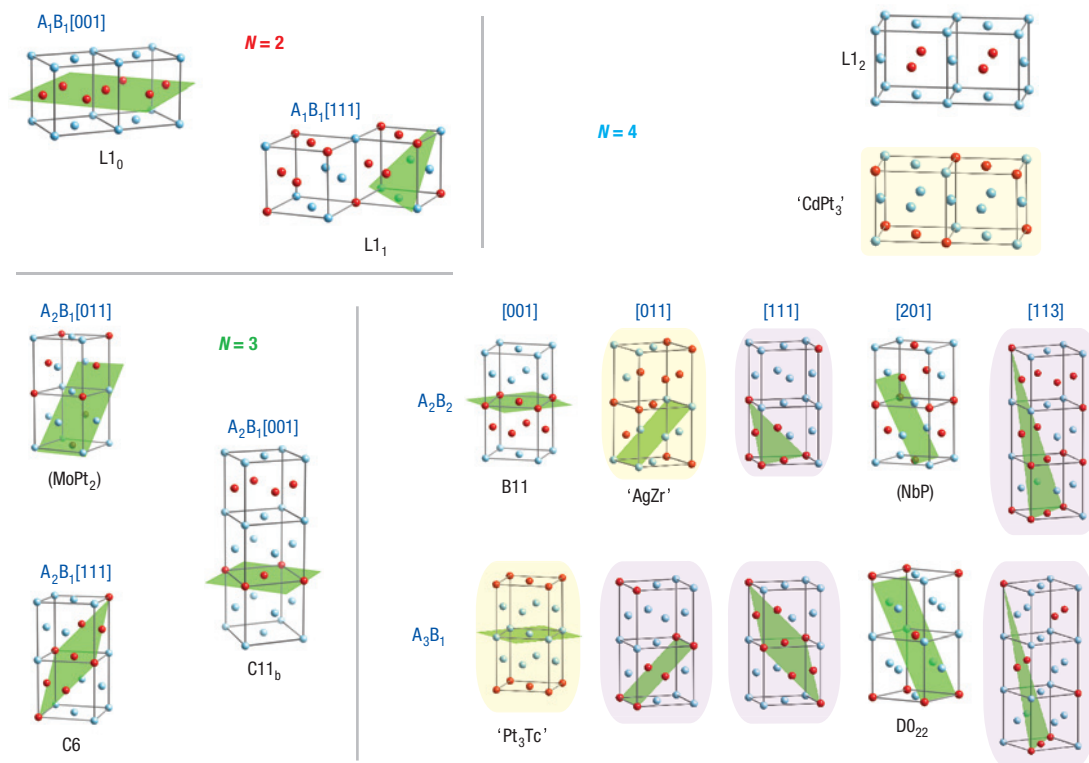


Figure 1 All 17 geometrically possible structures with four or fewer atoms. Of these 17 structures, 9 are well-known intermetallic structures. The other 8 have never been observed (shaded purple), but 3 have recently been predicted (shaded yellow). Structures are labelled by their common *Strukturbericht* names if assigned, otherwise by their common prototype in the case of NbP and MoPt₂ structures. Stacking sequences and directions are indicated for the 15 structures that can be so described. Predicted structures are labelled by quotes and named after their predicted prototype system.

kind will be more favourable. Structures whose geometry results in a 'random' distribution of like and unlike bonds will not be stable because chemistry dictates that either one or the other is energetically more favourable, for a given choice of atoms, but not both simultaneously.

The approach is best illustrated with an example. Imagine the following typical scenario. In an intermetallic compound, chemistry often favours unlike bonds between nearest-neighbour atoms but favours like bonds for next-nearest neighbours. In other words, unlike bonds are attractive (energy lowering) in the first 'shell' and like bonds are more energy lowering in the second shell. For the f.c.c. lattice, each atom has 12 nearest neighbours in the first shell and 6 second-nearest neighbours in the second shell. Because of geometrical 'frustration', it is impossible to arrange atoms to achieve more than 8 out of 12 A–B bonds in the first-nearest-neighbour shell. (See Fig. 2.) But it is possible to achieve 6 out of 6 A–A or B–B bonds in the second shell. It is interesting to note that the CuAu structure (L1₀), shown in the upper left in Fig. 1 and in Fig. 2, has the maximum possible number of nearest-neighbour unlike bonds (8/12) and next-nearest neighbour like bonds (6/6), and then to note how common that structure is—the L1₀ structure occurs in about 50 intermetallic compounds. Similarly, the L1₂ structure shown in the upper right of Fig. 1 has 6/12 nearest-neighbour unlike bonds and maximizes the number of next-nearest neighbour like bonds—and it occurs in about 200 compounds.

To summarize the discussion to this point: a geometric counting approach to enumerating possible unit cells, starting with the smallest first, suggests possible new crystal structures. Then, these can be further sorted according to bonding configurations of the

individual structures. Structures that have bonding configurations that deviate the most from a random configuration are the most likely to have a physical manifestation—irrespective of whether they favour like or unlike bonds at a particular bond distance.

This deviation from a purely random configuration can be quantified simply, as shown in Fig. 3. This figure shows the 'average bond type' for each bond length, for each of the 17 structures. Describing an unlike bond (A–B) by -1 and like bonds (A–A/B–B) by $+1$, we can sum over all the bonds in a structure and take the average at each bond length (nearest-neighbour, next-nearest neighbours and so on). With this description of bonds, a purely random configuration of atoms on the lattice will have a bond average of $(2x - 1)^2$, where x is the concentration of B atoms in an $A_{1-x}B_x$ compound. The deviation is simply the difference between the average bond type for a particular structure and the average bond type for a random configuration of the same concentration. In the upper left of Fig. 3, we can see that the nearest-neighbour bond average is $-1/3$ for the L1₀, the smallest possible value allowed by geometry. Similarly, the next-nearest neighbours bond average is $+1$, the maximum possible value and the maximum possible deviation from a random configuration (for which the bond average would be zero as $x = 1/2$).

Taking all of the relevant bond lengths in aggregate gives a single measure of a structure's deviation from the random case. The total bond average is just the absolute value of the deviations averaged over a range of bond lengths. Figure 4 shows the averaged deviation of each of the 17 structures taken over the first 10 bond lengths. This total deviation yields a simple measure of each structure's relative likelihood. The general result agrees strongly

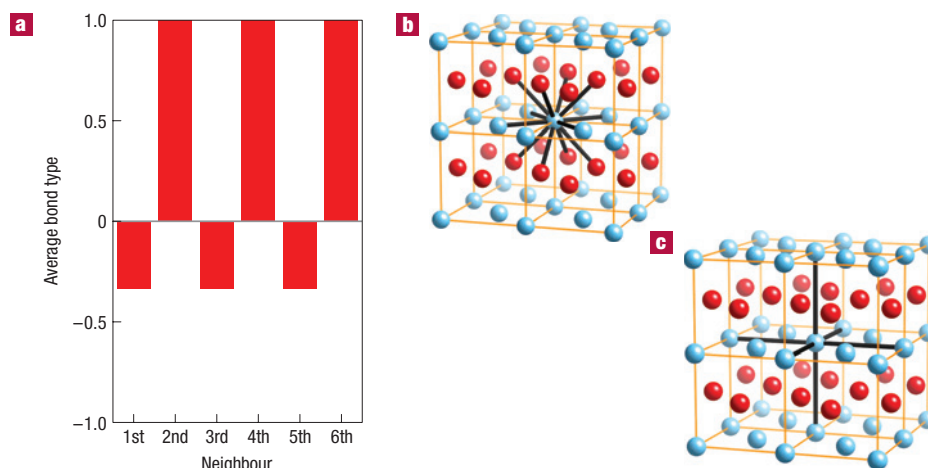


Figure 2 The average bond type for the $L1_0$ structure. **a**, The average bond type for the first 6 shortest bond lengths. **b**, At the nearest-neighbour distance, the central atom has 12 bonds forming 8 unlike bonds and 4 like bonds. As all atoms in this structure have the same nearest-neighbour environment, the central atom is representative of all atoms. Thus, the average nearest-neighbour bond type is $-8/12 + 4/12 = -1/3$. **c**, At the second-nearest-neighbour distance, the central atom has 6 bonds, all like. Again, as all atoms in this structure have the same environment, the average second-nearest-neighbour bond type is the same for each atom, so the global average is $+6/6 = +1$.

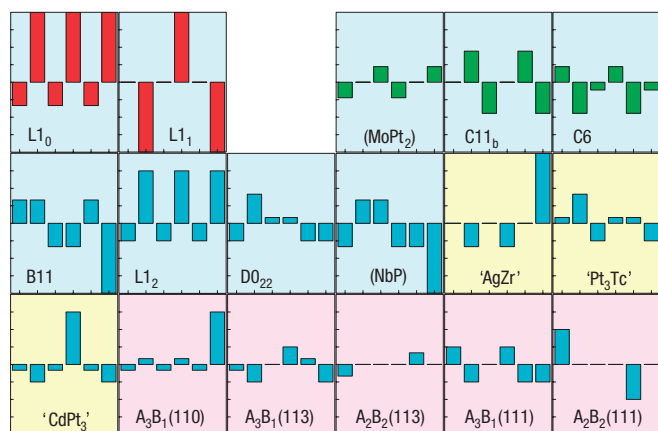


Figure 3 Deviations of bond averages. The horizontal axis of each plot corresponds to the first 6 nearest-neighbour distances. The vertical bars represent the deviation for each bond length. Negative values indicate a majority of unlike bonds, positive values like bonds. Structures that have been observed experimentally have light blue backgrounds, predicted but as-yet-unobserved structures yellow and never-observed or predicted structures have purple backgrounds.

with our observations—the likelihood measures of structures of known compounds are high, but low for unobserved structures. Even more compelling is the likelihood measures of predicted but as-yet-unobserved structures. They lie intermediate between the two extremes, suggesting that the predictions may actually be verified with experimental effort.

Two more interesting points emerge from examination of Fig. 4. One is the high relative likelihood of one structure that has not yet been observed or even predicted. The structure is an A_2B_2 [111]-stacked structure. This is a starting point for predicting or experimentally discovering a new structure. Also interesting is

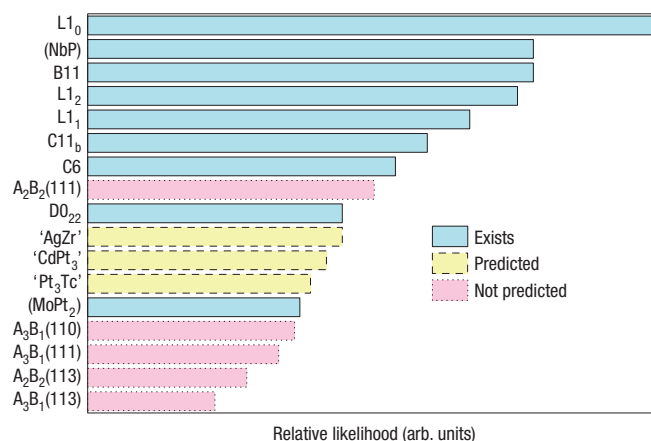


Figure 4 Likelihood measures (total bond averages) for each compound, ranked in descending order. The colour scheme is the same as that in Fig. 3. In general, structures with high likelihood measures correspond to those that have known physical manifestations, whereas structures that have never been observed or predicted to exist have the lowest measures. Predicted but as-yet-unobserved structures lie in between.

the 'CdPt₃' structure that was inferred based on a data mining technique⁷. A recent direct enumeration study concluded that among essentially all possible f.c.c.-based structures, this structure is indeed the stable one in at least two systems, Cd–Pt and Pd–Pt (ref. 20).

Two further examples illustrate the generic nature of this approach. Figure 5 shows the case of binary body-centred-cubic (b.c.c.)-based compounds. Again, the same general pattern emerges—observed structures have the highest likelihood measures, those that are not have the lowest; predicted structures lie in between. Next, consider Pt_8X_1 -type compounds²¹. About a dozen of these compounds have been observed but all have the same

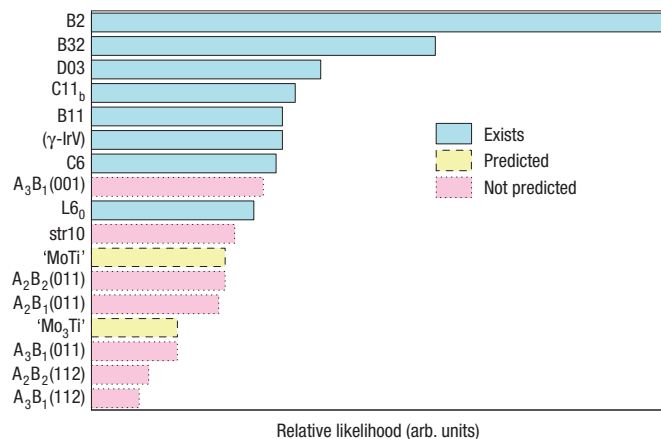


Figure 5 Likelihood measures for b.c.c.-based structures with four atoms per cell or fewer. The colour scheme is the same as that in Fig. 3. The general trend is the same as for the f.c.c.-based structures.

structure. Combinatorics reveals 14 possible f.c.c.-based structures that have 9 atoms/cell. Of these, the observed structure has the highest likelihood measure—it is the structure that deviates most from a random configuration, and it is this structure that appears over and over again in 8:1 intermetallic compounds.

The operating principle of this geometric approach is simply an extremum principle—the stability of a structure is the result of minimizing the energy (or free energy). Selecting structures with maximum deviation from a random configuration is synonymous with finding an extremum of the energy—doing so maximizes the number of energy-lowering or energy-raising bonds. In this way, the approach is conceptually parallel to that of the ‘interaction approach’ of Ducastelle²², Ceder and co-workers²³ and others^{24–26}, where the total energy can be expressed as a sum over short-range interactions. In that approach, the likely structures are predicted for all possible values of a few short-range interactions, whereas this geometric approach implicitly considers a larger number of interactions. Although the scope of the former is somewhat narrower, the two methods do make largely coincident predictions in the area where their range of applicability overlaps.

Whether a high-likelihood-index structure minimizes or maximizes the energy depends on the ‘chemistry’, the type of interactions, of a particular case. For one choice, a structure’s energy may be a minimum, but in others a maximum. This is evident in comparing the L1₀ and B11 structures in Figs 3 and 4—each has a high likelihood index, but the conditions under which L1₀’s energy will be minimized are nearly opposite those for B11 and vice versa. Thus, this approach finds new candidate structures in a chemistry-independent way.

This approach infers the possible existence of new structures on the basis of their geometric simplicity. The difference between modern-day heuristic methods and this approach is that the former predict new compounds, whereas this approach suggests new structures. In addition to the three examples given, the approach could be immediately applied to other important systems such as hexagonal close-packed (h.c.p.)-based binary compounds, polynary systems and cases of larger unit cells. Roughly speaking, there are about 1,500 binary intermetallic systems (most of which are f.c.c.-, b.c.c.- or h.c.p.-based) and about 50 times as many ternary systems. A comprehensive first-principle-based search of possible crystal structures for all these systems is not yet feasible but this approach provides a powerful starting point.

METHODS

RELATIVE LIKELIHOOD

Following is one final note about Figs 4 and 5. The computed total bond averages—the relative likelihood index—was computed by all bonds up to the tenth-nearest neighbour. This is an arbitrary cutoff—we could use a longer or a shorter cutoff. And, it could also be argued that long-range bonds should contribute less to the total energy of the crystal than the short-range bonds. In fact, only a finite number of bonds are unique in each structure, so a cutoff can be rigorously defined²⁷.

However, in reality, the important features of Figs 4 and 5 and the ensuing conclusions are unaltered by the choice of cutoff or whether or not shorter-ranged bonds are given heavier weight in computing the average. Thus, the results of this analysis are independent of the details of how the bond average is computed.

In the case of the Pt₈X₁ compounds, the small concentration of X atoms results in a more open structure that necessarily requires a longer cutoff to capture the essential chemistry. One structure has a likelihood index close to that of the physically observed structure. This other structure should probably be considered in the data mining approach and other ‘survey’-type ground-state searches.

STRUCTURE ENUMERATION

The details of the structure enumeration algorithm will be given in a forthcoming article but are outlined here. The first step generates all superlattices of a given parent lattice by enumerating all Hermite normal form matrices in order of ascending determinant sizes, as described in refs 18,19. The resulting superlattices are then reduced to a symmetrically inequivalent list using the point group of the parent lattice. Next, all possible atomic orderings, for each volume (determinant size), can be generated. This can be done using well-known combinatorial arguments by recognizing that there is an isomorphism between the quotient group of the superlattice and the direct sum of cyclic groups of the Smith normal form corresponding to each Hermite normal form matrix.

STRUCTURAL ASSOCIATIONS

Some of the structures shown in Fig. 1 can be assigned to different parent lattices depending on *c/a* ratios, internal coordinates and so on. For example, the physical prototypes of C6 and C11_b are h.c.p.- and b.c.c.-like respectively. But these are isotopic with the f.c.c. case and f.c.c.-like predictions have also been made.

Received 1 June 2007; accepted 9 October 2007; published 11 November 2007.

References

- Maddox, J. Crystals from first principles. *Nature* **335**, 201 (1988).
- Chaikin, P. Random thoughts. *Phys. Today* 8–9 (June 2007).
- Hume-Rothery, W. Researches on the nature, properties, and condition of formation of intermetallic compounds. *J. Inst. Met.* **35**, 319–325 (1926).
- Hume-Rothery, W. *Phase Stability in Metals and Alloys* 3–23 (McGraw Hill, New York, 1967).
- Miedema, A. R., de Boer, F. R. & de Chatel, P. F. Empirical description of the role of electronegativity in alloy formation. *J. Phys. F* **3**, 1558–1576 (1973).
- Pettifor, D. G. The structures of binary compounds. I. Phenomenological structure maps. *J. Phys. C* **19**, 285–313 (1986).
- Curtarolo, S., Morgan, D. & Ceder, G. Accuracy of ab initio methods in predicting the crystal structures of metals: review of 80 binary alloys. *Calphad* **29**, 163 (2005).
- Fischer, C. C., Tibbetts, K. J., Morgan, D. & Ceder, G. Predicting crystal structures by merging data mining with quantum mechanics. *Nature Mater.* **5**, 641–646 (2006).
- Sanchez, J. M., Ducastelle, F. & Gratias, D. Generalized cluster description of multicomponent systems. *Physica A* **128**, 334–350 (1984).
- de Fontaine, D. Cluster approach to order-disorder transformations in alloys. *Solid State Phys.* **47**, 33–176 (1994).
- Zunger, A. in *Statics and Dynamics of Alloy Phase Transitions* (eds Turchi, P. E. A. & Gonis, A.) 361–419 (NATO ASI Series, Ser. B, Plenum, New York, 1994).
- Daw, M. S. & Baskes, M. I. Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. *Phys. Rev. Lett.* **50**, 1285–1288 (1983).
- Daw, M. S. & Baskes, M. I. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443–6453 (1984).
- Sluiter, M. H. F. Some observed bcc, fcc, and hcp superstructures. *Phase Transit.* **80**, 299–309 (2007).
- Santorio, A. & Mighell, A. D. Properties of crystal lattices: The derivative lattices and their determination. *Acta Crystallogr. A* **28**, 284–287 (1972).
- Santorio, A. & Mighell, A. D. Coincidence-site lattices. *Acta Crystallogr. A* **29**, 169–175 (1973).

20. Hart, G. L. W. A new intermetallic prototype? Verifying new structure predictions in CdPt and PtPd, in *Bulletin of the March Meeting* (American Physical Society, 2007, <http://meetings.aps.org/link/BAPS.2007.MAR.W23.11>).
21. Pietrokowsky, P. Novel ordered phase, Pt₃Ti. *Nature* **206**, 291 (1965).
22. Ducastelle, F. *Order and Phase Stability in Alloys* (Cohesion and Structure, Vol. 3, North-Holland, Amsterdam, 1991).
23. Garbulsy, G. D., Tepesch, P. D. & Ceder, G. in *Materials Theory and Modeling* (eds Broughton, J., Bristowe, P. & Newsam, J.) 259–265 (Materials Research Society, Pittsburg, 1993).
24. Kanamori, J. & Kakehashi, Y. Conditions for the existence of ordered structure in binary alloy systems. *J. Physique* **38**, 274 (1977).
25. Allen, S. M. & Cahn, J. W. Ground state structures in ordered binary alloys with second neighbor interactions. *Acta Metallurg.* **20**, 423–433 (1972).
26. Sanchez, J. M. & de Fontaine, D. in *Structure and Bonding in Crystals* Vol. II (eds O'Keeffe, M. & Navrotsky, A.) 117–132 (Academic, New York, 1981).
27. Drautz, R. *Ab-initio Statistical Mechanics for Ordering and Segregation at the (110) Surface of Ni90%-Al*. Thesis, Max-Planck-Institut für Metallforschung (2003).

Acknowledgements

The author gratefully acknowledges financial support from the National Science Foundation through Grant Nos. DMR-0244183 and DMR-0650406. I would like to thank the following individuals for fruitful discussions: R. Drautz, S. Barthlein, S. Müller, M. Leone and B. Kolb. Input from V. Blum was particularly useful.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>