Table 7.1 Intrinsic vs. extrinsic dimensionality[a].

System	Extrinsic dimensionality	Intrinsic dimensionality
Au_8Pd_4	39	10.9
$Mg_{16}O_{16}$	99	11.6
$Mg_4N_4H_4$	39	32.5

[a]The numbers were produced using the molecular visualization toolkit STM4 [10] using the Grassberger–Procaccia algorithm with a suitably adapted Camastra sampling correction [11].

''explode'' the unit cell and we would not get any meaningful results. Therefore, we have to apply constrains to discard unfeasible solutions. The structure is considered unfeasible if:

1) The distance between any two atoms is smaller than threshold determined by user (e.g., there are no known bonds shorter than 0.5 Å). One can set different thresholds for different pairs of atom types separately; for example, the sum of correspondent atom radii.
2) One of the lattice vectors is too small. User can determine the threshold value; for example, it can be set to the diameter of the largest atom present in the system.
3) The angle between two lattice vectors is too small or the angle between the lattice vector and the diagonal of the parallelogram formed by other two lattice vectors is too small, see Figure 7.3. One can always choose the lattice vectors in such a way that the angle between any of them is in the $(60°, 120°)$ range. To do this, one can replace the longer vector (let's say $\mathbf{a}$) in a pair violating the constrain by $\mathbf{a}' = \mathbf{a} - ceil\left(|\mathbf{a} \cdot \mathbf{b}|/\|\mathbf{b}\|^2\right) \cdot sign(\mathbf{a} \cdot \mathbf{b}) \cdot \mathbf{b}$ see [20] for more details. Our experience shows that this procedure speeds up local optimization and improves the efficiency of the algorithm as a whole[2].

7.1.4
Initialization of the First Generation

Common way to create the first generation in evolutionary approach is uniform random sampling, aimed at achieving high diversity of the population. In fact, if you do not have any *a priory* information about how the possible optimal solution could look like, random sampling is the most reasonable way to do the unbiased

2) The 60–120 degree conditions are a simplified version of the cell constraints – more complete conditions are that a unit cell vector (and also cell diagonals) should not have projections onto other cell vectors that are greater by absolute value than half of the latter vector [20].
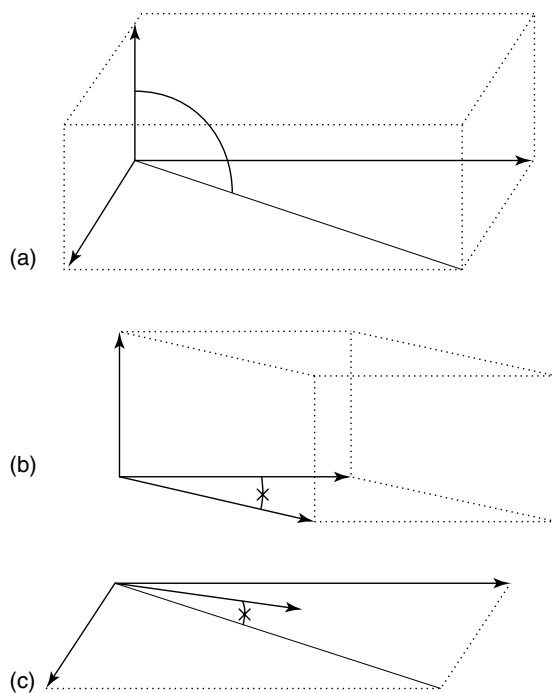
Figure 7.3 Cell angles constrains [20]. (a) All constrains satisfied, (b) angle between lattice vectors is too small, and (c) angle between lattice vector and the diagonal of the parallelogram formed by other two lattice vectors is too small.

initialization. You just pick up random points in the search space and check them for feasibility. Feasible solutions are added to the first generation. This process is repeated until the desired number of trial solutions is reached. In our case it means that we randomly choose the lattice vectors and then randomly drop the atoms into the unit cell. If some information about the optimum solution is known, for example, unit cell volume or lattice parameters, this can be used as constraints for optimization. In this case one would, for example, fix the lattice parameters and vary only atomic positions within the unit cell. User can also ''seed'' the first generation with the structures that seem reasonable (e.g., those known for similar compounds, or for the same compound at different conditions, or coming from previous structure prediction runs) and fill the rest of it with random ones.

High diversity of the first generation is the key to the success of the algorithm. If we do not have a structure in the basin of attraction (so-called funnel, see Figure 7.4) of the global minimum, then the probability to find that optimal solution can be low.

Random initialization, however, poses a problem relevant for large systems. When the number of atoms in the unit cell rises, randomly generated structures become more and more similar [15] from chemical point of view to each other