Embedding protein 3D-structures in a cubic lattice.
I. The basic algorithms.

Jacques Gabarro-Arpa

LBPA CNRS, Ecole Normale Supérieure de Cachan
61, Avenue du Président Wilson - 94235 Cachan cedex - France

e-mail: jga@grtran.org

Abstract

Realistic 3D-conformations of protein structures can be embedded in a cubic lattice using exclusively integer numbers, additions, subtractions and boolean operations.

1. Introduction

In previous papers [1–5] we have built a series of mathematical tools for studying the multidimensional molecular conformational space of biological macromolecules, with the aim of understanding the dynamical states of proteins by building a complete energy surface [6, 7].

An N-atom molecule has a $(N−1)^3$-dimensional conformational space (CS), the sheer complexity of this huge structure can be reduced to tractable dimensions by partitioning it with central hyperplanes into a finite set of cells, this amounts to discarding all knowledge about molecular conformations other than the cells that contain them.

In our approach [1], a set $H$ of $N_H = N \times (N−1)/2$ hyperplanes generates a partition in CS of $N!^3$ cells, on the other hand hyperplanes are oriented structures dividing the space into a + and a − half-spaces, thus points within a cell are characterized by a binary sequence of length $N_H$ enumerating the orientations with respect the hyperplane set. This binary sequence is all the information that remains from the molecular conformations.

Our choice of hyperplanes $\{H_{ij} : c_i − c_j = 0, \quad 0 ≤ i < j ≤ N − 1, \quad c \in \{x, y, z\}\}$[1], is such that the +/− hemispaces are the points with $c_i > c_j$ and $c_i < c_j$ respectively. This induces an order relation in the $x$, $y$ and $z$ coordinates of points in a cell

$$c_{\alpha_0} < c_{\alpha_1} < c_{\alpha_2} < ... < c_{\alpha_{N−2}} < c_{\alpha_{N−1}}$$

(1)

where $\{\alpha_0, \alpha_1, \alpha_2, ..., \alpha_{N−2}, \alpha_{N−1}\}$, a permutation of the sequence $\{0, 1, 2, ..., N − 2, N − 1\}$, is the dominance partition sequence (DPS)[1].

1That pass through the origin.

2A convention used here is that $c$ represents any of the cartesian coordinates $x$, $y$, $z$. 
The compactedness and hierarchical structure of the codes generated by partition sequences made possible the construction of a graph whose nodes are the cells in CS that are visited by the thermalized molecule with edges towards adjacent cells. This was the subject developed in previous works [2 - 5]. However interesting this result may be, it is of no practical use unless on top of it there is a method for calculating the energy of molecular conformations in a cell. With the mesoscopic force field approximations currently used in molecular simulations [8, 9], where atoms are represented as point-like structures, the only input to the Hamiltonian energy function are the interatomic distances calculated from 3D molecular conformations. In this framework the purpose of this work is twofold:

1. given a partition sequence, we want to calculate a fair sample of compatible 3D molecular conformations,

2. we want to encode the set of sampled conformations with a combinatorial structure so they can be more easily manipulated.

In the following sections are described the algorithms for doing this:

- In section 2 we build a complete set of lattice covalent bond segments, which are the basic building blocks: the whole molecular structure is built upon them.

- The DPSs can be seen as the lattice projections of a molecular structure where all intervals in each dimension are reduced to one lattice spacing (Fig. 5 of [1]), these have to be increased locally to obtain a realistic structure. In section 3 we build the partially ordered set of lattice intervals between bonded atoms, a structure needed for calculating the maximum and minimum expansion values of each interval, this gives a set of linear inequalities described in section 4.

- In section 5 it is shown how an inter-dependent system of inequalities can be made independent.

- In section 6 the form and structure of the system of linear inequalities is discussed in detail.

Figure 1: Stereoview of a pancreatic trypsin inhibitor protein (PTI) Cα-backbone molecular conformation (Table I), corresponding to the dominance partition sequences in Fig. 2.

To illustrate the algorithmic methods that are the subject of the present work, we have chosen as an example (Fig. 1 and Table I) the Cα-backbone of the pancreatic trypsin inhibitor protein [10], because it is a small protein molecule and the mathematical structures it generates are of moderate size, yet it has the complexity that can be found in longer molecules. Also the side chains have been put aside for the same reason: they would have made the contents of Figs. 2 and 3 almost unreadable.
2. The expanded lattice covalent bond segments set

The numbers in $x$, $y$ and $z$ dominance partition sequences can be regarded as the evenly spaced projections of $N$ points in a 3D cubic lattice, it is a particular form of embedding where the separation between consecutive projections of atoms in $x$, $y$ and $z$ has been shrunk to one lattice spacing. The aim of the present work is to expand this embedding so to obtain realistic molecular structures.

To do this we must restrict the most basic element of molecular structures: the covalent bond, to a finite set of coordinate values, such that with a suitable unit of length can be transformed to give integer values exclusively. These restricted bonds can still be useful for describing real molecular conformations if the minimum magnitude of vector differences is small enough. This can be done, for the example developed here ($PTI$ $C\alpha$-backbone), using empirical data sampled from molecular dynamics simulations [11], it requires the following steps

1. First we determine the dimensions of the lattice by taking as reference the mean bond length and its range of variation for bonded $C\alpha$ pairs, in our case this gives: $3.58\text{Å} < 3.86\text{Å} < 4.13\text{Å}$. We set arbitrarily the bond mean length to $20$ lattice units, which gives a lattice spacing of $0.19\text{Å}$. Thus, any segment between two lattice points with a length range between $3.58 \times 20/3.86$ and $4.13 \times 20/3.86$ is potentially a $C\alpha-C\alpha$ bond segment, and the set $B$ of valid lattice bond segments, modulo a lattice translation along the $x$, $y$ and $z$ axes, is the set of segments starting at the origin and ending in any lattice point that lies between two spheres of radius $3.58 \times 20/3.86$
2. Next we determine the range of variation for the bond angles, which is greater than that for the bond length and varies considerably along the Cα-chain. For each bond angle \( A_{\alpha_i, \alpha_{i+1}, \alpha_{i+2}} \) we determine two integer numbers : the floored minimum \( \lfloor \min(A_{\alpha_i, \alpha_{i+1}, \alpha_{i+2}}) \rfloor \) and the ceiled maximum \( \lceil \max(A_{\alpha_i, \alpha_{i+1}, \alpha_{i+2}}) \rceil \) respectively. These divide the interval between the absolute minimum and maximum values 71° - 167° in 64 subintervals

\[
\begin{align*}
71° &- 75° &- 79° &- 81° &- 82° &- 87° &- 90° &- 92° &- 93° &- 94° &- 95° &- 96° &- 97° &- 98° &- 99° &- 100° &- 101°. \\
104° &- 105° &- 106° &- 107° &- 108° &- 109° &- 110° &- 112° &- 113°. \\
114° &- 115° &- 116° &- 117° &- 118° &- 119° &- 120° &- 121° &- 124°. \\
125° &- 127° &- 129° &- 135° &- 136° &- 138° &- 139° &- 143° &- 144°. \\
147° &- 148° &- 149° &- 150° &- 151° &- 152° &- 153° &- 154° &- 155°. \\
156° &- 157° &- 159° &- 162° &- 163° &- 167°.
\end{align*}
\]

3. The dynamic values of each \( A_{\alpha_i, \alpha_{i+1}, \alpha_{i+2}} \) span a given range of intervals from (2), thus consecutive bonds \( B_{\alpha} \) and \( B_{\alpha+1} \) can only be assigned discrete bond segments that form an angle within the specific range.

In building realistic 3D-conformations from the DPSs by embedding these in a bigger lattice, the following problem arises: the intervals \( C_{\alpha_i} - C_{\alpha_{i+1}} \) between consecutive \( C_{\alpha} \)'s, for a given coordinate in Fig. 2, must be replaced by lattice intervals which are generally longer, so the excess lattice units must be distributed among the intermediate sequence intervals, such that the resulting lattice segments bonding \( C_{\alpha} \)'s are from the set of valid lattice bond segments described above.

To solve this problem the following steps are needed:

1. build from the DPSs the consecutive \( C_{\alpha} \) intervals poset (Fig. 3),
2. determine for each consecutive \( C_{\alpha} \) interval the maximum an minimum excess values,
3. make the linear inequalities in \( x, y \) and \( z \) independent of one another.

3. The consecutive \( C_{\alpha} \) intervals poset

Fig. 2 shows the DPSs for the PTI \( C_{\alpha} \)-backbone, it also shows some of the intervals between consecutive \( C_{\alpha} \)'s: \( T_{\alpha_1}^{c} \) a partial order relation can be defined for them. But first, we recall some basic definitions : let \( T_{\alpha_1}^{c} \) and \( T_{\alpha_2}^{c} \) be two \( T_{\alpha}^{c} \)'s spanning the DPS\(_{c}\) intervals \( \{ \sigma_{\alpha_1}^{cleft}, \sigma_{\alpha_1}^{cright} \} \) and \( \{ \sigma_{\alpha_2}^{cleft}, \sigma_{\alpha_2}^{cright} \} \)

**Definition 1** \( T_{\alpha_1}^{c} \) precedes \( T_{\alpha_2}^{c} \) or \( T_{\alpha_1}^{c} \prec T_{\alpha_2}^{c} \),
if \( T_{\alpha_1}^{c} \subset T_{\alpha_2}^{c} \) or equivalently \( \sigma_{\alpha_1}^{cleft} \geq \sigma_{\alpha_2}^{cleft} \) and \( \sigma_{\alpha_1}^{cright} \leq \sigma_{\alpha_2}^{cright} \).

**Definition 2** \( T_{\alpha_2}^{c} \) succeeds \( T_{\alpha_1}^{c} \) or \( T_{\alpha_2}^{c} \succ T_{\alpha_1}^{c} \).

**Definition 3** A maximal interval is not succeeded by any other interval.

**Definition 4** A minimal interval is not preceded by any other interval.

Fig. 2 shows the set of maximal intervals for DPS\(_{x}\), DPS\(_{y}\) and DPS\(_{z}\).

**Definition 5** A cover is a set of two intervals \( T_{\alpha_1}^{c} \prec T_{\alpha_2}^{c} \) with no \( T_{\alpha}^{c} \) such that \( T_{\alpha_1}^{c} \prec T_{\alpha}^{c} \prec T_{\alpha_2}^{c} \).

\(^3\)The following naming convention applies to any symbol refering to a bond interval \( C_{\alpha} - C_{\alpha+1} \): it bears only the smaller index.
Figure 3: Consecutive $C_\alpha$ intervals cover graph. Minimal/maximal intervals are at the bottom/top respectively, with succession going from bottom to top. For each interval

Fig. 3 displays a graphical representation of this partially ordered set (poset), where the nodes are the $I^\alpha_C$ set and the edge set consists of the pairs satisfying the cover relation. As we shall see below the poset structure allows to define the set of linear inequalities for determining the lattice bond segments.
4. Determining the bounds on excess values

The excess value of an interval $\mathcal{I}_{c}^{x}$ is the difference between its length on the DPS and on the extended lattice. In order to expand the DPS lattice we must determine first the bounds of excess values for every $\mathcal{I}_{c}^{x}$.

![Figure 4: Sequence of maximal interval $\mathcal{I}_{52}^{x}$ showing its minimal preceding intervals.](image)

An example will help to understand, we have in Fig. 4 a set of 5 connected $\mathcal{I}_{c}^{x}$s: $\mathcal{I}_{30}^{x}$, which is a maximal interval, and its minimal predecessors $\mathcal{I}_{30}^{x}$, $\mathcal{I}_{49}^{x}$, $\mathcal{I}_{54}^{x}$ and $\mathcal{I}_{53}^{x}$ (Fig. 3), they fill positions 8 to 20 in the $x$-sequence where the 12 minimal intervals between $C_{\alpha}$s have local excess variables $\chi_{5}^{x}$ to $\chi_{20}^{x}$ (Fig. 4), giving the local expansion value in the extended lattice. The following equations define the excess values

\[
X_{52}^{x} = \sum_{9 \leq \sigma \leq 20} \chi_{\sigma}^{x} - |\mathcal{I}_{52}^{x}| \quad \text{(where the last term is the c-sequence interval length)}
\]

\[
X_{30}^{x} = \sum_{10 \leq \sigma \leq 12} \chi_{\sigma}^{x} - |\mathcal{I}_{30}^{x}| \quad X_{49}^{x} = \sum_{11 \leq \sigma \leq 13} \chi_{\sigma}^{x} - |\mathcal{I}_{49}^{x}|
\]

\[
X_{54}^{x} = \sum_{15 \leq \sigma \leq 16} \chi_{\sigma}^{x} - |\mathcal{I}_{54}^{x}| \quad X_{53}^{x} = \sum_{17 \leq \sigma \leq 20} \chi_{\sigma}^{x} - |\mathcal{I}_{53}^{x}|
\]

also $X_{52}^{x}$ must be greater that the sum of the $X_{c}^{x}$ from preceding non-overlapping intervals

\[
X_{52}^{x} \geq X_{49}^{x} + X_{54}^{x} + X_{53}^{x} \quad X_{52}^{x} \geq X_{49}^{x} + X_{54}^{x} + X_{53}^{x}
\]

To build from (4) a complete system of linear inequalities allowing to calculate the $\chi_{c}^{x}$s for embedding the molecular system in the extended lattice, first we need to determine the bounds

\[
X_{min_{c}}^{x} \leq X_{c}^{x} \leq X_{max_{c}}^{x}
\]

By construction the maximum lattice bond segment length on any coordinate is 21, this gives for the extreme values of excess lattice units on any interval $\mathcal{I}_{c}^{x}$ the relation

\[
0 \leq |\mathcal{I}_{c}^{x}| + X_{c}^{x} \leq 21
\]

which settles the initial minimum and maximum bond lattice units for the c-coordinate to

\[
b_{c}^{min} = 0 \quad \text{and} \quad b_{c}^{max} = |\mathcal{I}_{c}^{x}| + 21
\]

respectively. Let $B_{c}^{b_{c}^{min}, b_{c}^{max}}$ be the set of all lattice bond segments $b$ such that $b_{c}^{min} \leq b_{c} \leq b_{c}^{max}$ for $c \in \{x, y, z\}$, then the set $B_{\mathcal{I}_{c}}$ of all the lattice bond segments that are within the bounds (7) is

\[
B_{\mathcal{I}_{c}} = B_{x}^{b_{x}^{min}, b_{x}^{max}} \cap B_{y}^{b_{y}^{min}, b_{y}^{max}} \cap B_{z}^{b_{z}^{min}, b_{z}^{max}}
\]

This operation may change the bounds (7), this is because the $b \in B_{\mathcal{I}_{c}}$ have a common origin but the points at the other extreme form a connected irregular cluster (see the example in Fig. 5): the bonds excluded by (8) may be the ones that contain the extremes of other coordinates. This gives a new set of bonds and the process has to be repeated until the bounds stabilize.
Figure 5: 2D example of a $B_{I_\alpha}$ set. The lattice bond segments ($b$) start at the origin, and end on any lattice point in the region bounded by the two spheres described in section 2.1. The only $b$ end points shown are those lying within the $x$ and $y$ bounds, or equivalently $b \in B_{I_\alpha}$.

As shown in the picture the set $B_{I_\alpha}$ can be decomposed into the minimal covering set of $n_p = 4$ rectangular subsets $P^0_{I_\alpha}$, $P^1_{I_\alpha}$, $P^2_{I_\alpha}$ and $P^3_{I_\alpha}$.

5. Making the inequalities independent

From the set of bounds (7) we can build the set of linear inequalities (using again the example from the previous section)

\[
\begin{align*}
X_{max}^x &\geq \sum_{9 \leq \sigma \leq 20} \chi^x_{\sigma} \geq X_{min}^x_{52} \\
X_{max}^x &\geq \sum_{10 \leq \sigma \leq 12} \chi^x_{\sigma} \geq X_{min}^x_{30} & X_{max}^z &\geq \sum_{11 \leq \sigma \leq 13} \chi^z_{\sigma} \geq X_{min}^z_{49} \\
X_{max}^y &\geq \sum_{15 \leq \sigma \leq 16} \chi^y_{\sigma} \geq X_{min}^y_{54} & X_{max}^z &\geq \sum_{17 \leq \sigma \leq 20} \chi^z_{\sigma} \geq X_{min}^z_{53}
\end{align*}
\]

(9)

There is a further problem to be taken into consideration: $X_{min}^c$ and $X_{max}^c$ are the $c$-coordinate bounds of the set $B_{I_\alpha}$, but, due to the non-uniform shape of $B_{I_\alpha}$, selecting one or more $c$-values in this interval while discarding the rest may change completely the bounds in the other coordinates. This induces an interdependence between inequalities (9) in $x$, $y$ and $z$, in which case solving the system becomes much more complex.

This problem can be avoided if the end points of bonds in $B_{I_\alpha}$ fill completely a lattice rectangular parallelopiped, in this case the choice of bounds in one coordinate leaves the others unchanged. Thus $B_{I_\alpha}$ has to be decomposed into a set of rectangular parallelopipeds $P_{I_\alpha}$

\[
B_{I_\alpha} = \bigcup_{0 < p \leq n_p} P^p_{I_\alpha}, \quad P^p_{I_\alpha} \in P_{I_\alpha}
\]

(10)

subject to the following conditions

1. there are no $P^p_{I_\alpha} \in P_{I_\alpha}$ and $P^{p_2}_{I_\alpha} \in P_{I_\alpha}$ such that $P^p_{I_\alpha} \subset P^{p_2}_{I_\alpha}$,

2. $n_p$ is minimal,

3. for $P_{I_\alpha}$ obeying conditions 1 and 2 and $P^{p_1}_{I_\alpha} \in P_{I_\alpha}$ there is no $P^{p_2}_{I_\alpha}$ such that $|P^{p_1}_{I_\alpha}| < |P^{p_2}_{I_\alpha}|$. 

In that case shrinking the bounds of a set \( P_p \) for any coordinate does not alter the bounds in the other dimensions and thus solutions to the inequalities can be found independently for each coordinate.

6. The structure of the solutions

The inequalities (9), for instance, can be rewritten as

\[
X_{max}^{a_2} \geq U_{9,20} \cdot x \geq X_{min}^{a_2} \quad \ldots
\]

where the \( U_{Ta} \)s are \((N - 1)\)-dimensional vectors of the form

\[
U_{Ta} = (0, ..., 0, 1, ..., 1, 0, ..., 0)
\]

with ones in the contiguous positions from \( \sigma^{cleft}_a \) to \( \sigma^{cright}_a \) and zeros everywhere else, and \( \chi^x \) is the vector

\[
\chi^x = (\chi_0^x, \ldots, \chi_9^x, \ldots, \chi_{20}^x, \ldots, \chi_{N-1}^x)
\]

Extending this notation to the whole set of inequalities for \( 0 \leq \alpha \leq N - 1 \) and \( x, y \) and \( z \), we have

\[
X_{max}^{a} \geq U^{a} \cdot x \geq X_{min}^{a} \quad \ldots
\]

The above set of inequalities define \( 2 \times (N - 1) \) affine half-spaces \( H_{min}^{a} \) and \( H_{max}^{a} \) whose intersection determines an H-polytope in \( CS^{c} \) [12, 13]. Hence, the vertices of this polytope are among the unique solutions of the \( 3 \times 2^{N-1} \) systems of equations

\[
X^{x} \cdot \chi^{x} = X_{lim}^{x} \quad U^{y} \cdot \chi^{y} = X_{lim}^{y} \quad U^{z} \cdot \chi^{z} = X_{lim}^{z} \quad 0 \leq \alpha \leq N - 1
\]

where \( X_{lim}^{c} \) can be either \( X_{max}^{c} \) or \( X_{min}^{c} \) and the \( \geq \) relation in (15) has been restricted to \( = \). Moreover, the matrices \( U^{c} \) with rows like (12) are called interval matrices, they belong to a very important class of matrices called: totally unimodular matrices [12]. These have the particularity that the determinant of any minor is either \(-1\), 0 or 1. This ensures that the vertices of the polytope are integer vectors (or lattice points), since solving (16) by applying the Cramer’s rule the denominator is always \(-1\) or \(1\). Thus, the solutions of (16) can be written

\[
\chi^{c} = U^{-1} \cdot X_{lim}^{c}
\]

where \( U^{-1} \) is the inverse of \( U^{c} \).

The V-polytope is the representation of the polytope by its set of vertices, these can be obtained from (17) by determining the combinations in \( X_{lim}^{c} \) compatible with (15). The solutions of the system of linear inequalities (15) can be generated from this set through convex combinations, as the three sets of inequalities are independent the general solution will be the product of the \( x, y \) and \( z \) polytopes.

The total unimodularity of matrix \( U^{c} \) also ensures that most combinatorial algorithms can be run in polynomial time.
7. Conclusion

The purpose of the line of work being developed here, is to show that molecular structures can be built and analysed with a fraction of the information (in our case less than $1/5$) that can be found in a typical PBD file.

This might seem a significant but modest quantitative difference, but qualitatively is more than that: discarding information results in the emergence of mathematical structures that were buried in the complexity of the data, which in turn can be encoded efficiently by them. Using combinatorics a great number of molecular conformations can be dealt simultaneously, thus overcoming the barrier that computations have to be performed on the basis of one conformation at a time.

The algorithmic method developed before [1−5] serves two purposes

1. As an amplifier: by codifying data sampled in computer simulations into discrete geometrical structures, these can be combined to generate an estimate of the volume occupied by a molecule in its conformational space.

2. As a molecular 3D-structure compressor: it is possible to translate basic features of molecular 3D-structures into a binary code, which in turn can be very efficiently amalgamated into ternary sequences that encode great numbers of cells from CS. The information on the whole CS volume can be cast into a file compatible with desktop memory size.

The present work is the first one of a third and last step: the development of combinatorial methods for calculating the energy of structures from cells in $CS$.

Here we have developed the basic algorithms for this: realistic discrete protein conformations can be built and embedded in a cubic lattice, using a table of discrete bond segments and, more important, these conformations can be encoded into combinatorial structures.

However many issues still remain unexplored:

- The possible combinations of $P^p_{\lambda\alpha}$ s from (10) is a huge set, efficient sampling methods should be developed.
- The V-polytope should be better characterized.
- The present formalism should be extended to take into account sets of adjacent cells.
- Last of all inter-atomic distances should also be encoded into combinatorial structures.

These will be dealt in forthcoming works.
8. Appendix

Table 1. Lattice coordinates of the PTI \( C_\alpha \)-backbone from Fig. 1.

Column \( \alpha \) : \( C_\alpha \) number.

Columns \( x_\alpha \ y_\alpha \ z_\alpha \) : \( C_\alpha \) coordinates.

Columns \( b_x \ b_y \ b_z \) : bond vector between \( C_{\alpha-1} \) and \( C_\alpha \).

| \( \alpha \) | \( x_\alpha \) | \( y_\alpha \) | \( z_\alpha \) | \( x_\alpha \) | \( y_\alpha \) | \( z_\alpha \) | \( b_x \) | \( b_y \) | \( b_z \) |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0           | 0           | 0           | 0           | 29          | -34         | 49          | -30         | 0           | 19          | 7           |
| 1           | 19          | 7           | 1           | 30          | -33         | 66          | -40         | 1           | 17          | -10         |
| 2           | 30          | 5           | -16         | 31          | -25         | 84          | -38         | 8           | 18          | 2           |
| 3           | 31          | 26          | -19         | 32          | -9          | 94          | -44         | 16          | 10          | -6          |
| 4           | 12          | 26          | -26         | 33          | -1          | 113         | -41         | 8           | 19          | 3           |
| 5           | 19          | 19          | -43         | 34          | 14          | 111         | -29         | 15          | -2          | 12          |
| 6           | 28          | 35          | -49         | 35          | 30          | 123         | -29         | 16          | 12          | 0           |
| 7           | 19          | 52          | -53         | 36          | 38          | 117         | -12         | 8           | -6          | 17          |
| 8           | 27          | 68          | -45         | 37          | 55          | 106         | -15         | 17          | -11         | -3          |
| 9           | 29          | 86          | -53         | 38          | 64          | 91          | -25         | 9           | -15         | -10         |
| 10          | 28          | 106         | -48         | 39          | 50          | 80          | -34         | -14         | -11         | -9          |
| 11          | 47          | 105         | -41         | 40          | 44          | 61          | -30         | -6          | -19         | 4           |
| 12          | 60          | 120         | -46         | 41          | 36          | 55          | -13         | -8          | -6          | 17          |
| 13          | 54          | 131         | -31         | 42          | 17          | 51          | -19         | -19         | -4          | -6          |
| 14          | 40          | 145         | -33         | 43          | 12          | 69          | -11         | -5          | 18          | 8           |
| 15          | 24          | 141         | -21         | 44          | -3          | 68          | 2           | -15         | -1          | 13          |
| 16          | 6           | 137         | -29         | 45          | -14         | 83          | 12          | -11         | 15          | 10          |
| 17          | 2           | 126         | -15         | 46          | -32         | 75          | 11          | -18         | -8          | -1          |
| 18          | -15         | 111         | -21         | 47          | -44         | 65          | -1          | -12         | -10         | -12         |
| 19          | -8           | 93          | -14         | 48          | -50         | 50          | 10          | -6          | -15         | 11          |
| 20          | -16          | 76          | -19         | 49          | -33         | 46          | 18          | 17          | -4          | 8           |
| 21          | -7           | 60          | -28         | 50          | -24         | 47          | 0           | 9           | 1           | -18         |
| 22          | -13          | 42          | -32         | 51          | -38         | 32          | -4          | -14         | -15         | -4          |
| 23          | -15          | 41          | -52         | 52          | -34         | 23          | 13          | 4           | -9          | 17          |
| 24          | -10          | 22          | -57         | 53          | -15         | 22          | 9           | 19          | -1          | -4          |
| 25          | -18          | 25          | -75         | 54          | -17         | 18          | -11         | -2          | -4          | -20         |
| 26          | -36          | 29          | -67         | 55          | -24         | -1          | -8          | -7          | -19         | 3           |
| 27          | -35          | 17          | -51         | 56          | -36         | -12         | 3           | -12         | -11         | 11          |
| 28          | -40          | 30          | -37         | 57          | -52         | -15         | -9          | -16         | -3          | -12         |
References

[1] A central partition of molecular conformational space. I. Basic structures.
   J. Gabarro-Arpa, Comp. Biol. & Chem. 27, 153-159 (2003).

[2] A central partition of molecular conformational space. II. Embedding 3D structures.
   J. Gabarro-Arpa, Proceedings of the 26th Annual International Conference of the IEEE EMBS,
   San Francisco, 3007-3010 (2004).

[3] A central partition of molecular conformational space. III. Combinatorial determination
   of the volume spanned by a molecular system in conformational space.
   J. Gabarro-Arpa, J. Math. Chem. 42, 691-706 (2006).

[4] A central partition of molecular conformational space. IV. Extracting information
   from the graph of cells.
   J. Gabarro-Arpa, J. Math. Chem. 44, 872-883 (2006).

[5] A central partition of molecular conformational space. V. The Hypergraph of 3D
   Partition Sequences.
   J. Gabarro-Arpa
   arXiv:0812.2844 (2008).

[6] Energy landscapes.
   D.J. Wales, Cambridge University Press, ISBN 0-521-81415-4, (2003).

[7] Dynamic personalities of proteins.
   K. Henzler-Wildman, D. Kern, Nature 450, 964-972 (2007).

[8] All-atom empirical potential for molecular modeling and dynamics studies of proteins.
   A.D. MacKerell Jr., et al., J. Phys. Chem. B 102, 3586–3616 (1998).

[9] Biomolecular simulations: recent developments in force fields, simulations of enzyme
   catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent inter-
   actions.
   W. Wang, O. Donini, C.M. Reyes, P.A. Kollman
   Annu. Rev. Biophys. Biomol. Struct. 30, 211–243 (2001).

[10] The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen
    and its complexes with inhibitors.
    M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber
    Acta Crystallogr. Sect. B 39, 480–490 (1983).

[11] Clustering of a molecular dynamics trajectory with a Hamming distance.
    J. Gabarro-Arpa, R. Revilla, Comput. Chem. 24, 693-698 (2000).

[12] Theory of Linear and Integer Programming.
    A. Schrijver, John Wiley & sons, ISBN 0-471-98232-6, pp. 155-156 (1998).

[13] Lattice points and lattice polytopes.
    A. Barvinok, Handbook of Discrete and Computational Geometry, CRC Press, ISBN 0-8493-8524-
    5, pp. 133-152 (1997).