We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



28

Evolutionary Algorithms in Crystal Structure Analysis

Attilio Immirzi[1], Consiglia Tedesco[2] and Loredana Erra[3]
[1,2]*Dept. of Chemistry, University of Salerno*
[3]*European Synchrotron Radiation Facility*
[1,2]*Italy*
[3]*France*

1. Introduction

Evolutionary algorithms (EA, see for a general introduction Holland, 1975; Goldberg, 1989; Davis, 1991; Back, 1996) are "probabilistic" optimization procedures. Differently from the "deterministic" ones (yielding an unique and reproducibile solution for assigned input data), EA are somewhat aleatory and not reproducible. Such methods are part of the heuristics (Polya, 1971). To be pedantic the term algorithm should indicate some *sequence of mathematical operations producing a foreseeable result if applied to definite input data*. There is a resounding oxymoron in the terms *evolutionary algorithm*. This contradictory language is however diffuse; as this one is not an erudite essay of epistemology, we prefer to be tolerant and fly over on semantic questions. Really evolutionary methods have aleatory routes and also aleatory conclusions in many cases (e.g. stock exchange estimate, weather forecast, etc.); in some cases instead (it happens in crystallography), unique and certain results can be reached in spite of the multiplicity of routes travelled.

Multi-parameter optimization problems are encountered in many fields: industrial process planning, financial investments, environment control, hurricane evolution, and many others (Weise, 2009). We shall however not consider the general aspects of EA but rather to treat about a specific argument: the study of chemical structures based on X-ray diffraction. In this field, as shown later, there is a lot of *a priori* and useful information so that the aleatority of the procedure can be considerably attenuated. Evolutionary methods in crystallography are better formulated as *Constraint Satisfaction Problems* (CSP, see Ionita et al., 2010).

Evolutionary methods have a future in crystallography, particularly in solving hard problems. The structural elucidation from X-ray powder diffraction data is among these (section 4.1). However, procedures are not consolidated yet, there are numerous alternatives to be considered for which the common sense and the experience, rather than the theory, could have a role.

This article will not treat the EA in the crystallographic context in its generality; rather it will summarize some ideas and strategies which yielded a specific procedure: it is a possible route, not the only possible one. These procedures are now included in a general-purpose computer program with numerous options for finding and refining crystal structures termed TRY (Immirzi, 2007b). In its actual structure the EA procedure needs that many parameters

are assigned by the user. Assigning them in a more automatic fashion could no doubt improve the program versatility.
The effectiveness of EA in solving crystal structures by diffraction methods is improved if one makes systematically use of the so called *internal coordinates* (i.c.) for describing structures, instead of the more common *cartesian coordinates*. The i.c. were first considered, in diffraction analysis, by Arnott & Wonacott (1966). A computer program performing least-square refinement of i.c. (LALS, Linked Atom Least Square) was first issued (Smith & Arnott, 1978) and later updated repeatedly (the last edition, WINLALS , was published by Okada et al., 2003). LALS has been used mainly in studying polymers. Also Tadokoro (1975) discussed the use of i.c. in polymer case.
Both Arnott (and disciples) and Tadokoro did however neglect a very important complication: the mutual independency (non-redundancy) of i.c., being of great importance in all crystallogaphic applications, and in particular in EA, as will be shown later.
Now a procedure for carrying-out a molecular building *always using non-redundant i.c.* has been devised (Immirzi 2007a,b); it is very simple, practical, and perfectly analytical, runs thoroughly, and eliminates the necessity of using exotic instruments in matrix manipulations (pseudoinversion, diagonalization, etc.; note that the mathematic instruments employed are those of Newton, Gauss, and Lagrange).

2. Crystalline matter and X-ray diffraction

Characterization of chemicals, both natural and synthetic, is done always looking at the matter at the atomic scale. In 19th century, atoms were speculative entities and scientists had not instruments for ascertaining their shape or position. Nevertheless, in the middle of century, some of them (Kekulé, Le Bel, van't Hoff and others), with evident attitude versus the heuristics (without awareness of course as heuristics did not exist yet) were persuaded about molecules (proposed many years before by Avogadro finding much skepticism); they begun, with much imagination and using logical arguments, to give them definite shapes, for instance the tetrahedral shape to methane and the hexagonal shape to benzene. Many of these intuitions were demostrated later be perfectly exact. Today to ascertain the geometrical structure of molecules is not more a work of imagination: it is an exact science. Structure is a "property" of any pure substance, defined and reproducible, like colour, density, melting temperature, etc.[1]
The discovery of X-ray diffraction by crystals by Laue in 1912 made possible to ascertain the structure. In the crystalline state, necessarily solid, atoms are ordered: there is a small portion of space, a parallelepiped, generally oblique, termed the *unit cell*, which repeats itself by translation in three directions and generates the whole solid. If the disposition of atoms in the cell is determined, the whole crystal is determined. Frequently there is also symmetry: only a fraction of the unit cell is independent; the remainder if given by appropriate orthogonal transforms. Edges and interaxial angles of the unit cell are collectively termed *lattice constants*.
Just order makes possible X-ray diffraction of crystal, whose study permits to guess the matter at the atomic scale; resolution is of the order of 0.02 Å or better; consider that the separation of bonded atoms is 1-2 Å. The size of molecules studied by X-ray diffraction was initially modest; now studying 50-100 atom molecules is a common matter.

[1] There are indeed exceptions: some substances exhibit more than one structure (*polymorphism*). Crystalline minerals are frequently polymorphic.

3. Molecules and molecular crystals

Molecules are finite assemblies of atoms joined by strong forces (chemical bonds). In molecular crystals molecules are orderly assembled through well weaker non-bonding forces (van der Waals forces, dipole-dipole forces, hydrogen bonds, etc.).[2]

Thanks to the repetition, crystals diffract X-ray radiation: when a beam of monochromatic radiation strikes a crystalline sample many beams emerge and their intensities can be one-by-one measured. The intensity of the diffracted beams is related to the molecular structure (see below).[3]

In absence of symmetry, if V is the volume of the unit cell, λ the radiation wavelength, and $2\vartheta_m$ the highest angular deviation, there are altogether $3.35(V/\lambda^3)\sin^3\vartheta_m$ different diffracted beams (reflections). The above relationship indicates that the number of reflections raises as $\sin\vartheta_m/\lambda$ increases. The latter is an important parameter: the higher is $\sin\vartheta_m/\lambda$ the higher is the chance of success in a structural analysis, and the higher is the accuracy of the result. Consider however that the intensity of the reflections decreases as ϑ angle increases and that $\sin\vartheta_m/\lambda$ may be limited by natural circumstances: there are materials which are intrinsecally weakly diffracting; their structural study cannot be done accurately.

4. Structure and diffraction

It is possible to obtain crystal structures from X-ray diffraction data thanks to the mathematical relationship between atomic positions and the intensities of the reflections. The latter, as numerous as the volume of the unit cell is large (see above), are singled out by three integer numbers: the Bragg indices (h,k,ℓ). Each is deviated from the incident direction by a characteristic angle 2ϑ, an analytical function of the Bragg indices and of the lattice constants. One must intercept these beams and measure intensity.

There are nowadays very sophisticated instruments (*diffractometers*), controlled by computers, allowing, with a minimal human intervention and in a short time, a complete characterization of a crystal: measurement of the lattice constants, diagnosis of symmetry, localization and measurement of all reflections with their Bragg indices h,k,ℓ. If the unit cell contains N atoms, and x_j,y_j,z_j are the atomic coordinates (referred to the unit cell), the complex quantities

$$F(h,k,\ell)=\sum_{j=1}^{N} f_j \exp[2\pi i(h\frac{x_j}{a}+k\frac{y_j}{b}+\ell\frac{z_j}{c})] \tag{1}$$

are termed *structure factors*. The f_j (atomic factors) are known real quantities which depend on the chemical nature of the j-th atom and on $\sin\vartheta/\lambda$. The larger is the atomic number the higher are f_j. In every cases f_j decreases as $\sin\vartheta/\lambda$ increases.

The above relationship applies at 0 K; at higher temperature F's are somewhat reduced because of the *thermal vibration*. There are of course as many $F(h,k,\ell)$ as reflections.

Now the squared moduli of $F(h,k,\ell)$ should be orderly proportional to the measured diffraction intensities $I(h,k,\ell)$. This occurs, of course, when the x_j,y_j,z_j coordinates are the true ones. To get the unknown crystal structure one must find all the x_j,y_j,z_j rendering the

[2] Not all crystalline materials are molecular . There also ionic crystals and *extended covalent structures*.

[3] The deviation of diffracted beams from the incident direction obeys to the Bragg's law. Geometry is quite similar to the case of a reflecting mirror: diffracted beams can be considered as "reflected" by the sample. For this reason they are frequently termed *reflections*

$|F(h,k,\ell)|^2$ as close as possible to the $I(h,k,\ell)$. Introducing the residual

$$\chi^2 = \sum_{hk\ell} [I(h,k,\ell) - |F(h,k,\ell)|^2]^2 \tag{2}$$

one can assume that the solution is the one rendering χ^2 a minimum. For this reason to find a structure is to solve a *global optimization problem*. The above χ^2, divided by the sum of $F^2(h,k,\ell)$ is the common R_2 index, widely used as *fitness function*.

4.1 Single-crystals and polycrystalline samples

X-ray diffraction studies use either *single-crystal samples* or *polycrystalline samples* (powder, fiber). In the second case things are more difficult because all diffracted beams having a given 2ϑ are overlapped. While the diffraction pattern of a single crystal is a three-dimensional function $I(h,k,\ell)$ rich of information, the powder diffraction is an uni-dimensional function $I(2\vartheta)$ with poor information.

Structural studies based on powder diffraction data meet the difficulty of measuring the diffracted intensities one-by-one, singled out by the Bragg indices, in presence of overlap. The difficulty can be overcome employing deconvolution techniques (see e.g. Harris, 1998) or, alternatively, renouncing to the separation of reflections and considering the quasi-continous function $I(2\vartheta)$ (Rietveld, 1967; Rietveld, 1969; Young, 1995) exploiting the so called *full-pattern powder profile analysys*.

In this case, the measured diffraction intensities $I(2\vartheta_i)$ are compared with the computed ones given by

$$I_{calcd,i} = \sum_k F_k^2 \, \Omega(2\vartheta_i - 2\vartheta_k) \tag{3}$$

where the sum is extended to *all the reflections whose Bragg* $2\vartheta_k$ *fall near to the current* $2\vartheta_i$ *point*. In equation (3) $\Omega(2\vartheta_i - 2\vartheta_k)$ is the *peak function* having a maximum for $2\vartheta_i = 2\vartheta_k$. In full-pattern profile analysis the residual χ^2 considered is given by the expression

$$\chi^2 = \sum_{2\vartheta_i} [I(2\vartheta_i) - I_{calcd,i}]^2$$

and the most used fitness index is $R_{wp} = \chi^2 / \sum I^2(2\vartheta)$

Just because EA techniques need a reduced number of data they are attractive in studies based on powder diffraction, particularly when the full-pattern profile analysis is performed. For recent studies see Harris (1998), Feng (2006), Hanson (2007), Oganov (2006).

When a structure has been solved (at a coarse level) a second problem arises: to refine atomic positions. This is a *local optimization problem*. Once again the goal is to find the minimum of χ^2 varying systematically atomic positions *in the vicinity of the initial coarse values* and adding some parameters for expressing the thermal vibration.

While the local optimization is done using deterministic algorithms (typically a least-square refinement) and proceeds, in most cases, without problems, the global optimizazion, even at coarse level, is much more exacting.

The remainder of this chapter concerns only global optimization problems. In addition we shall not treat the traditional solution methods but only some based on evolutionary algorithms.

5. The three-dimensional space and the problem space

Solving a structure of N atoms means finding N positions in the 3D space. Unfortunately there are no algorithms for finding positions one by one (unless rather simple cases occurring in inorganic chemistry). Atomic positions must be find simultaneously.
An alternative way to define the terms of problem is to refer to another space, the so called *problem space*, having $3N$ dimensions instead of three. To catch a structure means to locate a *single point* in this hyper-space. Simple to say, but very difficult to do. As explained later the EA follow this crazy idea.
Small-medium size problems, are solved using deterministic algorithms, provided that the available diffraction measurements are numerous, accurate, and complete (i.e. including all the intensities with the diffraction angles ϑ lower than the assigned $2\vartheta_m$). For these methods, not discussed in this article, there are many excellent textbooks (e.g. Giacovazzo et al., 2002; Stout & Jensen, 1995).[4] The EA are indicated instead in difficult cases: limited $2\vartheta_m$, poor quality of data, low number of reflections.
One could ingenuously believe that all structural problems can be solved moving the representative point systematically in the whole problem-space and examining all solutions. Atomic coordinates are however real and continuous quantities; so points are infinite. Well, even though finite (and coarse) intervals are considered to render the problem-space discontinuous, one obtains astronomic numbers also in the simple cases (see later). The systematic exploration of the whole problem-space is not a practical tool, unless some tricks are adopted. Possible strategies are:

i) to reduce substantially the number of variables adopting other kinds of coordinates rather than crystallographic (see section 6) and choosing coordinate systems for which a good fraction of variables are predictable;

ii) to render the problem-space discontinuous;

iii) to reduce as much as possible the range of variation for each variable.

We shall suppose that the crystal under study is a molecular crystal with known composition (*chemical formula*) and known interatomic connection (*structural formula*); consider that the chemical formula is obtained by chemical analysis; structural formula is the result of a number of physical observations (e.g. NMR, IR, UV spectroscopy) and chemical observations (relations of the unknown substance to other already known).
Using diffraction techniques one can establish size, shape, and symmetry of the unit cell. From the experimental density of crystals one obtains the atom content of the unit cell. Using the diffractometers one obtains an appropriate number of diffracted intensities.

6. Coordinates

Structures are commonly described by using the crystallographic coordinates (c.c.) to define atom positions. The c.c. (collectively indicated with p_i) are $3N$ if atoms are N. This choice is the most natural and also practical because the simple analytical relationship with the structure factors $F(h,k,\ell)$, see equation (1).

[4] Among deterministic methods we include the so called direct methods, actually the most used. Also direct methods are somewhat probabilistc and aleatory, in modest amount however. Also using direct methods the risk of finding false solutions exists.

As a matter of fact, infinite alternative systems for describing structures are conceivable: if new coordinates q_i are introduced, so that one can transform from p_i to q_i and viceversa *using analytic and biunivocal relationships*, the q_i can be used as alternative stuctural variables. The number of q_i must be, of course, again $3N$.[5]
Which are the alternative useful descriptions among the infinite possible ones? To reply it is important to point out that atomic positions are not fully impredictable. The systematic study of molecular crystals did demonstrate that atomic positions are not random quantities. Any structural hypothesis must obey to the following rules:

i) atoms must be appropriately separated among them; for each atom pair distance is anything but random: if the two atoms are chemically bound, atom separation is a bond-length (b-l) and must be close to the sum of the atomic radii (see e.g. Cotton, 1999); if not bound and spaced by three bonds or more, the distance must be higher than the sum of the so called van der Waals radii (Bondi, 1964); the same applies when the two atoms belong to different molecules (packing distances).

ii) the bond-angles (b-a) must be close to the "canonical" values prescribed by the rules of orbital hybridation (see e.g. Cotton, 1999): angles close to 109.5° on carbon atoms with sp^3 hybridation, close to 120° on carbon atoms with sp^2 hybridation, etc. The examination of "molecular models" suggest in much cases angles in rather restricted intervals.

iii) the molecular conformation must obey to the rules of stereochemistry: aromatic rings must have D_{6d} symmetry, the sequences C–C=C–C have to be planar, the torsion-angles (t-a) about single bonds must have values close to those typical of ethane-like molecules (-60, 60, 180°), etc. These limitations can be thus "rigid" in some cases (double bonds), "flexible" in other cases (single bonds).

The above restrictions for b-l, b-a, and t-a are of simple mathematical formulation if the crystallographic coordinates p_i are employed; there is however the disadvantage that the restrictions (*constraints*) apply *not to the p_i themselves, but to a number of mathematical functions of the p_i*. Indeed there are procedures for doing minimizations in presence of constraints: the Lagrange method (Goldstein, 1980) with the drawback that only a few constraints can be accounted for, while in molecular building there are numerous. Instead, introducing alternative coordinates q_i *chosen with cleverness*, it is possible to impose the above constraints not on the functions but directly on the q_i. If so happens one simply removes the q_i from the list of variables.

7. The internal coordinates. Eyring algorithm

It is customary studying complicate problem by means of the so called *internal coordinates* (i.c.), used in performing molecular building in various other contexts (e.g. spectroscopy, theoretical chemistry). The pioneer work (in chemistry) was that of Eyring (1932) who devised the procedure later called *Z-matrix* . Really the idea is well older; it was skilfully described by Lagrange (1796) who coined the term *generalized coordinates*. Lagrange studied not molecules but machines; things are rather similar after all: Lagrange's machines are made by rigid objects

[5] $3N$ coordinates are necessary for describing the *crystal structure*. If only the *molecular structure* is of interest, the coordinates are $3N - 6$ being six the rigid body coordinates (Goldstein 1980).

connected to each other with limited freedom of movement; in molecules the rigid objects are the chemical bonds and movements are limited to torsion and (in small amount) to bending.[6]
Eyring's algorithm is simple, provided that there are neither rings nor polyhedra. The idea is to build molecules first defining three atoms, (`C1` , `C2` , and `C3` in Fig. 1) then inserting the other atoms stepwise as function of three i.c. in each step: a b-l b, a b-a τ and a t-a ϑ. Of course, in each step one must specify which atoms are to be considered.
Three i.c. are needed for starting a molecular building, viz. two b-l (b_1, b_2) and a b-a (τ_3). It is customary to put the 1st atom on the origin, the 2nd atom along x axis, the 3rd atom in the x, y plane with $y > 0$.
Fig. 1 shows how the building starts (`chbe` command, green atoms) and the Eyring construction (red atom). The construction of the 5th atom (blue) is discussed later.

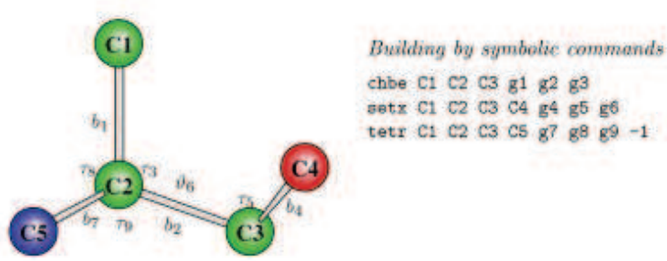


Fig. 1. Building a 5 atoms molecule

The Eyring's machinery for attaching a new atom after a given sequence A, B, C, is as follows: one computes the vector $\mathbf{v}$ with components $b\cos\tau$, $b\sin\tau\cos\vartheta$, $b\sin\tau\sin\vartheta$, the vector product $\mathbf{u} = (B - A) \times (C - B)$, and the orthogonal transformation matrix $\mathbf{T}$ aligning A and B points along z' and making $y'_C = x'_C = 0$; then adds to C atom the product $\mathbf{T} \cdot \mathbf{v}$. The analytical expression of $\mathbf{T}$ matrix is:

$$\mathbf{T} = \begin{vmatrix} x_B - x_A & x_C - x_B & u_x \\ y_B - y_A & y_C - y_B & u_y \\ z_B - z_A & z_C - z_B & u_z \end{vmatrix}$$

with each column normalized to length 1.
Besides the Eyring's construction (`setx`) needing three variables b, τ, ϑ, another construction is practical when atoms are inserted on tertiary carbon atoms *of known chirality*; this construction (`tetr`) makes use of two b-a τ_1, τ_2 (instead of one b-a and one t-a); the advantage is the restricted range of the b-a compared with the wide range of the t-a. The above 5-atom skeleton (Fig. 1) has been built just with 3 constructions: `chbe` (for atoms C1–C2–C3, in green), `setx` (for atom C4, in red), and `tetr` (for atom C5, in blue).
In Fig. 1 the three constructions are indicated "symbolically" according to a conventional syntax (see later).
Of course six other variables must be added to the molecular i.c. for defining the actual position and orientation of the molecule in the unit cell: three translations and three rotation angles. Rototranslation i.c. can be less than six because of crystal symmetry.

[6] This representation applies when bonds are considered rigid and bond-angles semirigid . This situation applies in studying molecular crystals (at a coarse level). In other contexts (e.g. spectroscopy) things are different.

7.1 Redundant coordinates

A serious drawback encountered in building molecules using Eyring method is the possible redundancy of the i.c. In simpler words *they might be not independent*. While independent i.c. can be varied each in turn preserving the *stuctural formula*, this does not apply in case of redundancy. When molecular building is performed within random search and genetic procedures, the preservation of the structural formula is of basic importance. Also regarding reduncancy there is nothing new after Lagrange; he said all the necessary: the *generalized coordinates* are exactly the same thing.
As illustrated below, non-redundancy in molecular building can be fulfilled by selecting the $3N-6$ i.c. of a molecule with shrewdness among b-l, b-a, t-a, and *also bending angles if necessary* (see later). Bending angles become necessary in two cases: cyclic molecules and polyhedric molecules.

7.2 Cyclic molecules

A possible solution overcoming redundancy in cyclic molecules is shown in Fig. 2 considering the case of cyclohexane.
Since the g_i must be, at a molecular level, $3N-6$, considering N bond-lengths, only other $2N-6$ g_i must be assigned, necessarily, angular. The method adopted (Immirzi, 2007a,b) integrates Eyring's procedure for building atoms C1, C2, ...C5 (9 g_i are employed altogether, 4 b-l, 3 b-a, and 2 t-a) with a new machinery for building atom C6 (proposed by Goto and Osawa, 1989) using two b-l and one angle only, a so called *bending angle* (φ); with this construction the molecule is bent about a line crossing two atoms separated by one bond only. Such construction (termed `flap`) has been added to the above ones. The machinery for computing C6 as a function of φ and the two b-l C1–C6 (g_{11}) and C5–C6 (g_{12}) is very simple (see Immirzi, 2005a).
The construction shown in Fig. 2, termed `flap` resolves the problem *for rings of any dimension* and also every polyciclic molecule (e.g. decaline, steroids, etc.) and resolves also intricate multicyclic molecules like norbornane, pinene, spirocompounds, etc.

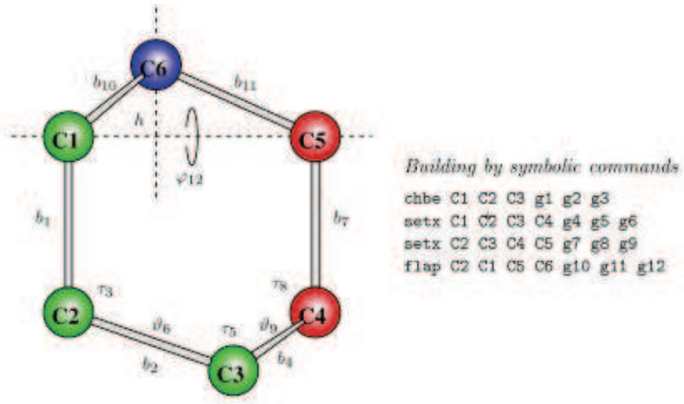


Fig. 2. Building a ring without redundancy, cyclohexane C_6H_6, (the blue atom is computed by a `flap` instruction).

7.3 Polyhedric molecules

A possible solution overcoming redundancy in polyhedric molecules is shown in Fig. 3 considering the simplest polyhedron: the cubane C_8H_8, represented as a bare carbon atom

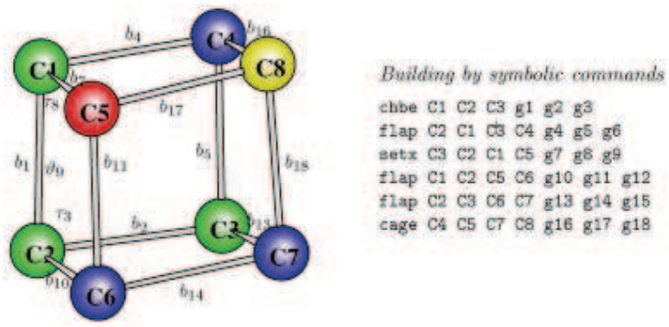


Fig. 3. Building a polyhedric molecule without redundancy (cubane C_8H_8). The green atoms are computed by `chbe`, the red atom by `setx`, the blue atoms are computed by `flap`, and the yellow atom by `cage`.

skeleton. Remember that the g_i must be, at the molecular level, $3 \times 8 - 6 = 18$ and 12 g_i are b-l. In order to build without redundancy one must use *only 6 angles*. Building the first 7 atoms can be done as indicated symbolically in the figure. After the 5th command one atom is lacking, C8, and *all the 6 angles have been used* besides 9 over 12 bond-lengths. To complete the molecule one can find x_8, y_8, z_8 imposing C8 has assigned distances b16, b17, b18 from points C4, C5, C7. This is a simple (and classical) problem of 2nd degree having an unique solution provided that the polyhedron is convex. The command `cage` computes the lacking atom. Note that `cage` allows modeling also giant polyhedra like the fullerenes.

8. Other constructions and symbolic building

Molecular and crystal building through internal coordinates is not used only in performing structural analysis by EA. It can be used instead in all conventional procedures. In particular for doing interactive modeling, trial-and-error calculation, and at last for doing the least-square refinement of structures based on the i.c. themselves as optimized variables. In the latter case the non-redundancy of the variables plays a crucial role since only in absence of redundancy normal matrices are always non-singular.[7]

For these reasons all the mentioned constructions have been programmed devising an unique subroutine (termed `LAGR` , an homage to Lagrange, the authentic, and ignored, discoverer of the i.c.) performing the whole construction at each call. To simplify the input data preparation `LAGR` has been structured so that the various constructions invoked are specified symbolically according to a conventional syntax with one line of data for each construction step. Examples of this syntax are indicated in the figures and in the input data for procedure validation (Tables 1 and 2).

Besides the mentioned four basic constructions needed for building molecules of any kind (`chbe`, `setx`, `tetr`, `flap`, and `cage`, `tetr` is really not indispensable) there are numerous other constructions, not strictly necessary but very useful. Among them there are `phen` (building phenyl groups), `metl` (building a CH_3 group) and others. `LAGR` subroutine includes also the commands necessary for placing the built molecule in the unit cell, for doing orthogonal transforms, for referring the molecule to its inertial axes, and others. A clever

[7] A persuasive argument is that normal matrix needed for carrying out the least-square procedure is computed evaluating the first-order partial derivatives of structure factors vs. the variables. Such derivatives are computed wrong if the variables are not strictly independent to each other.

usage of these commands brings to a further reduction of the effective number of i.c. by imposition of *local symmetry*, i.e. the identity of structurally independent i.c. but chemically indistinguishable. At last a number of special commands were implemented for the case of linear polymers: to create regular helices and chains with glide-planes, to orient the chain parallel to the crystallographic *c* edge, etc.

9. General layout of a EA based procedure. Metaheuristic approach

The evolutionary (or genetic) algorithms for resolving global optimization problems are based on the idea of finding the "true" solution starting from a number of more or less arbitrary solutions and performing combinations of the variables analogous to the ones occurring in the cellular reproduction.

There is an analogy between chromosomes and *molecular structures*. As a chromosome can be described by a sequence of genes, a molecular structure can be described by a sequence of structural variables. For these reasons the terms are used as synonyms.

In order to limit the aleatority of the procedure it is practical to follow the so called *metaheuristic approach* (Weise, 2009) assuming that the g_i vary as multiple of a small increments. With this contrivance the i.c. (real and continuous quantities) become integer numbers whose upper limit is limited by the range of excursion of the g_i themselves. "Solutions" can binary-coded using a finite number of bits. Consider anyway that the number of points in the problem-space remains astronomic; for instance a 10-variable structure codified in 60-bits may assume $2^{60} \approx 10^{18}$ values. If, for simplicity, a 5-gene case is considered, and 7, 6, 3, 4, 6 bits are dedicated to the 5 genes respectively, the binary code will be (braces groups the bits of given g):

$$\underbrace{b_6b_5b_4b_3b_2b_1b_0}_{g_1}\quad\underbrace{b_5b_4b_3b_2b_1b_0}_{g_2}\quad\underbrace{b_2b_1b_0}_{g_3}\quad\underbrace{b_3b_2b_1b_0}_{g_4}\quad\underbrace{b_5b_4b_3b_2b_1b_0}_{g_5}$$

9.1 Planning the complete structural building

It is propedeutical to write-up a complete schema for the structural building based on the symbolic language.[8] This important step is anything but "automatic"; rather it requires much attention and cleverness. One must assign the non-redundant i.c. and prepare the full list of the symbolic building commands which, using the aforementioned subroutine `LAGR`, will provide the full structural building as function of the current i.c. This schema serves not only for carrying out the EA procedure, but for *all the numerous options of TRY program* (calculation of structure factor, model adjustment, crystal packing, least-square refinement, geometrical computations, etc.). Among the symbolic instructions there is one indispensable for carrying out the random search in the problem space, termed `xcon`, and discussed in details later.

9.2 Binary encoding

Now the user must: i) decide which g_i must be kept fixed and which varied (it is customary to exclude the bond-lengths and, possibly, also the bond-angles); ii) assign to each variable g_i a convenient increment Δg_i (the g_i will be varied by multiple of Δg_i), iii) assign to each

[8] These schemas are familiar to people using the Eyring's mechanism for molecular building. In the present case things are a little bit more difficult for the variety of constructions, which are necessary to avoid redundancy.

variable g_i how many bit (m_i) are to be used in binary encoding. Δg_i and m_i determine the range of excursion of the g_i (range it is exactly $\Delta g_i \times 2^{m_i}$).
It is evident that some common sense must be exerted in assigning Δg_i and m_i with a compromise between fine steps in exploring the problem-space and wide spanning intervals. For instance, for a bond angle on a carbon atom one could assign $\Delta = 1.0^\circ$ and $m_i = 4$; the angle will span a sufficient interval of 16°. Torsion angles (about single bonds), have instead higher uncertainity; giving them e.g. 5 bits and $\Delta g_i = 3^\circ$ will permit a span of 64°. A large number of bits, e.g. 7 or 8, must be assigned to the molecular rotation angles: using $m = 7$ and $\Delta = 2.8^\circ$ will ensure a 360° span; with $m = 8$ the same is fulfilled for finer step: 1.4°. Using common sense is no doubt necessary; but also experience turns out to be useful.
In this mechanism the initial value of each variable g_i° (to be defined in the input data) plays a role, since the values assumed by g_i are ($k = 1, 2 \ldots 2^m$)

$$g_i^\circ,\ g_i^\circ - \Delta g_i,\ g_i^\circ + \Delta g_i,\ g_i^\circ - 2\Delta g_i,\ g_i^\circ + 2\Delta g_i,\ g_i^\circ - 3\Delta g_i, \ldots$$

Note however that the g_i° are not critical quantities, provided they are internal to the span intervals.

10. Creating the initial population. Constraints

The first step of an evolutionary procedure is the creation of a *population of tentative structures* by means of a Montecarlo method (using e.g. pseudo-random numbers).[9]
One must assign the size of the population desired (e.g. 100-300 items) and the selection criteria (see below). Size and selection criteria control, in a very unforeseeable manner, the duration of the search, from few minutes to days. Anyway the search can be in any moment interrupted and restarted with new parameters, either creating a new list of solutions or queuing to the existing one.
If K is the number of g_j searched (K is of course well lower than $3N$), the random search is performed generating, for each tentative structure, K integer random numbers each in the range $0 \div 2^{m_j} - 1$ to obtain a random combination of g_i. By using the cited subroutine `LAGR`, the corresponding model is created. It is evident that most of these models will be unfeasible. There are two alternatives: i) to retain all solutions, attaching to each solution a proper "penalty function", ii) to apply some selection criteria and reject immediately unfeasible items. The latter has been chosen in our procedure to avoid endless populations. Much experience is anyway necessary for calibrating the rejection parameters, actually under user control.

10.1 Selection criteria. The `xcon` command

The computer program implemented in our laboratory includes the following selection criteria (*filters*):

1) the 1st filter, most incisive and mandatory, is the one checking the connectivity of the trial structure versus the known connectivity (remember that the chemical formula is supposed known!). We define as *connectivity* a group of eight integer codes (`connv`) to be inserted

[9] True random numbers should be preferable. Some hardware devices are produced, but we do not have direct experience. Such devices should be of course *very fast* as billions random numbers are necessary.

in the `xcon` command: the first is the number of atom-pairs separated by one bonds, the second the same number for two bonds, etc. up to 8 bonds.[10]

2) the 2nd filter eliminates structures with unfeasible bond angles. Non bonded pairs must have distances exceeding more than `sepn2` (atoms separated by two bonds) or `sepn3` (atoms separated by three or more bonds) the sum of covalent radii. Unrealistic structures with acute bond angles can so be eliminated. (In few cases, eg. cyclobutane, acute angles occur). The quantities `sepn2` and `sepn3` are supplied in the command `xcon`.

3) the 3rd filter checks the connections of the trial molecule with the neighbour (in molecular crystals no connections should be present; in linear polymer, instead, two connections take place); the number of external connections (`linkno`) is also assigned in `xcon` command. Connection are illegal if distance is lower than the sum of covalent radii augmented by `tolnk`, also supplied in `xcon`.

4) the 4th filter consists in computing the lattice energy, again rejection occurs if energy exceeds the assigned threshold, chosen by the user at the beginning of search.

5) the 5th filter is based on the R_2 index defined as the residual χ^2 (equation 2) divided by the sum of intensities. the tentative solution is rejected if R_2 exceeds an assigned value. Also for R_2 the threshold is chosen by the user.

Parameters `sepn2`, `sepn3`, `tolnk`, `linkno` and `connv` are defined in command `xcon`. Note that the same filters apply also in the subsequent *breeding* (see later).
It is evident the importance of assigning the above values *cum grano salis*, avoiding both excess or lack of severity. Indeed the ratio accepted-structures / generated-structures leans to be low (e.g. 10^{-5}); the conformational freedom of the molecule has of course a critical role.

10.2 Optimizing of the initial population

We have introduced (as an optional) the *local optimization* of the random solutions found with the above procedure. Each random solution is a point in the (discontinuous) problem-space. The idea is to examine the nearest points in the problem-space and find points possibly more promising, based e.g. on R_2 index. In a K-dimension problem space points are $3^K - 1$, a value prohibitively high if K is large.
On belief that the local optimizazion is a good idea (experience seems to confirm) we have exploited a trick: to assign a reasonable value to the number of neighbouring points P to be examined, say 500-1000, and select, at random, P points among the $3^K - 1$ ones.[11] With this contrivance the duration of an optimization step is reasonable.

11. Genetic combinations

The heart of the genetic procedure is the systematic combination (*breeding*) of the solutions belonging to the initial population following the rules of genetics. In this context two kinds of

[10] Establishing the connectivity codes is rather bothering, but the problem is easily resolved using TRY. One considers a dummy unit cell without symmetry and with large lattice constants. If the molecule is well constructed and displayed on the monitor the eight connection codes will appear on the top of the screen. Keep note and insert in `xcon`.

[11] For each integer p running from 1 to P one computes a real random number in the range 0-1, multiply it by 3^K and truncate. The point to be considered is given by the *digits of the product above written in the basis 3*. If desired 5 or 7 points can be considered instead 3.

combinations are considered: the *cross-over* and the *mutation*. A cross-over (*binary combination*) consists of an interchange of a selected gene belonging to the two mating items; a mutation (*unary combination*) is the change of a single bit (chosen at random) belonging to the bit-string. Considering a 9-gene example and a cross-over between genes g_2, g_7 one obtains:

Parent sequence (father):	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9
Parent sequence (mother):	g'_1	g'_2	g'_3	g'_4	g'_5	g'_6	g'_7	g'_8	g'_9
Child sequence:	g_1	g'_7	g_3	g_4	g_5	g_6	g'_2	g_8	g_9
Mutated sequence:	g_1	g'_7	g_3	g^*_4	g_5	g_6	g'_2	g_8	g_9

the symbol $*$ has been used for a mutation (the g_4 has been chosen by chance). In each case a new sequence is born for each crossover, two if also a mutation takes place. If the crossover takes place for all the pairs of K genes, $K \times (K+1)/2$ new strings occur; twice as many if also mutation is done.

Of course if two feasible structures are coupled, it is not warranted that child structures are meaningful. The above described "filtering" could be used again to eliminate unfeasible structures. At present the structures having R_2 higher than an assigned limit or a lattice energy higher than an assigned limit are eliminated. The limits, of course, should be fixed using experience. The new selected structures, anyway, are added to the initial ones.

The crossover can be done either considering a single gene, selected at random, or considering all genes; in the last case there are K combinations, for each pair of mated structures. When the genetic combination is ended the population of solution is extended by an amount depending on the used filters. The new population will be again sorted using an appropriate figure like the R_2 index.

Examining the sorted list one observes, frequently, that *the first solutions happen to be very similar to each other*. That is no doubt an indication of success. If this is not the case a good idea is to repeat the breeding as many times as it is necessary. If however the breeding has resulted ineffectual in finding new solutions, it is better to repeat the whole process in more appropriate conditions. The exploration of a wide fraction of the problem-space is no doubt essential for the success.

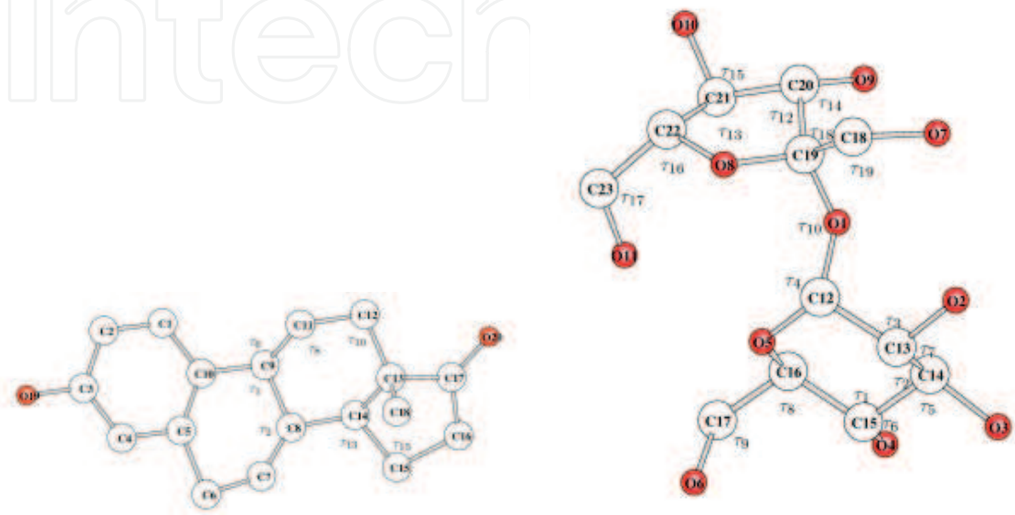
12. Procedure validation

For testing purposes two already reported structures (single-crystal X-ray diffraction studies, molecular models are shown in Fig. 4) have been considered using published data. Only the reflections having $d > 1.2$ Å have been considered to show that the method needs few reflections and that resolution can be modest.

The first test concerns the steroid equilin (a 20-atom molecule, ignoring hydrogen atoms) with restricted conformational freedom (Sawicki et al. 1999), the second test considers sucrose, a 23-atom molecule with high conformational freedom (Hynes et al., 1991). The latter was already considered in the first publication describing the method (Immirzi et al., 2008); the analysis has been repeated however in more challenging conditions. The input data for doing the two tests are given in Table 1 and Table 2; many comments are added for reader's convenience.

All computations were done using the program TRY (Immirzi, 2007) and assuming fixed bond-lengths in both cases, also fixed bond-angles in the case of equilin. The computing scheme is the simplest one: no hydrogen atoms, isotropic thermal vibration (an unique B_{iso}), unitary weight factors. In both cases the EA procedure consists of three phases: i) formation of

the initial population, ii) improvement of solutions looking at adjacent points, iii) breeding by cross-over and mutation. The reliability of the resulting structures has been tested performing some cycles of full-matrix least-square refinement optimizing the same i.c. considered in the EA phase. The resulting R indices are, of course, not competitive with the ones published because of the crude simplicity of the followed computation scheme and because of the reduced number of reflections. The LS convergency is anyway excellent.



(a) Equilin (b) Sucrose

Fig. 4. Molecular models of the two test molecules: equilin and sucrose. Bond angles (τ_n) are shown. Torsion angles (ϑ_n) are listed in Tables 1 and 3.

The following scheme summarizes the essential data for the two tests.

	equilin	sucrose
reflections considered ($d > 1.2$ Å)	715	589
number of non-hydrogen atoms	20	23
number of internal coordinates	60	68
varied internal coordinates	15	28
bits used in binary encoding	75	123
random structures generated	50×10^3	141×10^6
initial population	300	150
lowest R_2 index after search	0.43	0.74
lowest packing energy	-3Kcal	-11Kcal
search duration	3 min.	29 hr.
mated structures	20	30
resulting R_2 after breeding	0.39	0.43

Note how the formation of initial population is performed in a very short time in the former case (75-bit encoding), and longer time in the latter (123 bit encoding).

```
Equilin by Sawicki 1999
6.5429  9.0345  23.894  0  0  0  !  lattice constants
C=C  O=O                          !  atomic species
4 0.72080 56.55 4  4  6           !  3 non-operative codes + crystal extension
-x+1/2, -y, z+1/2,                !  symmetry (P2_12_12_1 space group)
-x, y+1/2, -z+1/2, *
!  here begin lines processed by LAGR subroutine
xcon  0.10 0.10 0.10 0  23  36  39  30  22  16  12  7
!  1st ring (aromatic)  C1-C2-C3-C4-C5-C10 (conventional geometry)
chbe  C1  C2  C3   1.40  1.40  1.20
setx  C1  C2  C3   C4    1.40  1.20  0.0
setx  C2  C3  C4   C5    1.40  1.20  0.0
setx  C3  C4  C5   C10   1.40  1.20  0.0
!  2nd ring (cyclohexene) C6-C7-C8-C9
setx  C2  C1  C10  C9   1.52 1.20 1.80
setx  C1  C10 C9   C8   1.52 g1  g2
setx  C10 C9  C8   C7   1.33 g3  g4
flap  C8  C7  C5   C6   1.52 1.52 g5
!  3rd ring:  (cyclohexane)
setx  C1  C10 C9   C11  1.54 g6   g7
setx  C10 C9  C11  C12  1.54 g8   g9
setx  C9  C11 C12  C13  1.54 g10  g11
flap  C9  C8  C13  C14  1.54 1.49 g12
!  4th ring (cyclopentane)
setx  C9  C8  C14  C15  1.54 g13  g14
setx  C8  C14 C15  C16  1.54 g15  g16
flap  C15 C16 C13  C17  1.52 1.52 g17
tert  C12 C14 C17  C13  C18  1.54
!  O atoms
setx  C5  C4  C3   O19  1.36 1.20  1.80
trig  C13 C17 C16  O20  1.22 0
iner  C1  Q31  20   0   !  refer the 20-atom molecule to inertial axes
rtax  1   C1  20  g18   !  Rx rotation
rtax  2   C1  20  g19   !  Ry rotation
rtax  3   C1  20  g20   !  Rz rotation
move  C1  20  g21 g22 g23  0  !  translation
end
1.0800     1.6900     1.2000     0.0800     1.7000  !  starting values
1.0900    -0.6400     1.0900     1.7800     1.0600  !  for i.c.
0.5450     1.2700     1.2200     1.7700     1.0200  !
1.6600     1.7200    -0.9521    -0.5334    -0.5628
2.0146     3.5320     9.8812
*
*
180 -90 180  50  1.0 1  !  screen projection and scale
5.2  5.2  !  B-iso      !  thermal parameters
0.0  0.0  !  B-33       !  anisotropic component of ditto
!  the following 15 lines define the g(i) kept variable in EA
!  and the binary-encoding scheme:  no.  of bit and step; span intervals
!  (sp) are also given as comments
 2  4  0.028  0  0  !  sp for C1-C10-C9-C8   t-a =45deg
 4  4  0.028  0  0  !  sp for C10-C9-C8-C7   t-a =45deg
 5  4  0.028  0  0  !  sp for C8-C7-C5-C6    t-a =45deg
 7  4  0.028  0  0  !  sp for C1-C10-C9-C11  t-a =45deg
 9  4  0.028  0  0  !  sp for C10-C9-C11-C12 t-a =45deg
11  4  0.028  0  0  !  sp for C9-C11-C12-C13 t-a = 45deg
14  4  0.028  0  0  !  sp for C9-C8-C14-C15  t-a = 45deg
16  4  0.028  0  0  !  sp for C8-C14-C15-C16 t-a = 45deg
17  4  0.028  0  0  !  sp for C15-C16-C13-C17 flap =45deg
18  7  0.028  0  0  !  sp for Rx rotation = 360deg
19  7  0.028  0  0  !  sp for Ry rotation = 360deg
20  7  0.014  0  0  !  sp for Rz rotation = 360deg
21  5  0.200  0  0  !  sp for Tx translation = 6 A
22  6  0.200  0  0  !  sp for Ty translation = 12 A
23  7  0.200  0  0  * !  sp for Tz transl.  = 25 A
0 1.0 10.0 0.1 1.0 1.0 1.0 1.0 0.8 1.0 0.0
EQUILIN.DAT ! file containing the F2hkl
```

Table 1. Input data for equilin test

```
Sucrose by Hynes (1991)
10.8631  8.7044  7.7624  90 102.94 90  !  lattice constants
C=C  O=O  H=H                          !  atomic species
8 1.5418 56.55  4  4  4                !  3 dummy codes and cell extension
-x, 1/2+y, -z, *                       !  symmetry (P2_1 space group)
!  here begin lines processed by LAGR subroutine
xcon  0.30  0.30 0.40  0 24 36 43 36 32 30 26 17 !  connections
chbe C16  C15  C14       1.521   1.521   g1    !  build piranose
setx C16  C15  C14  C13  1.521   g2      g21
setx C15  C14  C13  C12  1.521   g3      g22
flap C15  C16  C12  O5   1.427   1.427   g23
setx C16  O5   C12  O1   1.427   g4      g24   !  O atom bridge to furanose
setx C16  C15  C14  O3   1.427   g5      g25   !  O3
setx C13  C14  C15  O4   1.427   g6      g26   !  O4
setx C15  C14  C13  O2   1.427   g7      g27   !  O2
setx C14  C15  C16  C17  1.521   g8      g28   !  lateral CH2OH group
setx C15  C16  C17  O6   1.427   g9      g29
setx C13  C12  O1   C19  1.427   g10     g30   !  build furanose
setx C12  O1   C19  C20  1.521   g11     g31
setx O1   C19  C20  C21  1.521   g12     g32
setx C19  C20  C21  C22  1.521   g13     g33
flap C21  C22  O19  O8   1.427   1.411   g34
setx O8   C19  C20  O9   1.427   g14     g35   !  O lateral to furanose
setx C19  C20  C21  O10  1.427   g15     g36   !  O lateral to furanose
setx C19  O8   C22  C23  1.521   g16     g37   !  lateral CH2OH
setx O8   C22  C23  O11  1.427   g17     g38
setx C21  C20  C19  C18  1.521   g18     g39   !  lateral CH2O
setx C20  C19  C18  O7   1.427   g19     g40
iner C1   Q50  23   0  !  compute the inertial axes and orient accordingly
rtax 1    C1   23   g41 !  overall Rx rotation
rtax 2    C1   23   g42 !  overall Ry rotation
rtax 3    C1   23   g43 !  overall Rz rotation
move C1 23 g44 0 g45 0  !  translation
end                                    !  i.c.  starting values follow
   1.1072   1.0725  1.1207   1.1010  1.0771   1.1249   1.0982   1.1219   1.1124
   1.1381   1.0818  1.0225   1.0285  1.1583   1.1181   1.0999   1.1284   1.1479
   1.1137   0.0000 -0.5680   0.5729  1.2994  -0.6819 -1.7724  -1.7409   1.7851
   1.7344 -0.6413   1.3017   1.6018 -0.8762  -0.3520  1.7662   1.5806  -1.5559
  -1.3286  0.7047   1.4875   0.7208  0.7006  -0.1187   0.5964  2.4590    3.1670
*
-90  0  0  50  1.0  1    !  molecular orientation, and image scale factor
1.3  1.3  1.3            !  Isotropic thermal parameters
0.0  0.0  0.0            !  anisotropic component of ditto
!  binary-encoding scheme for g(i):  no.  of bit and step; span intervals  (sp) after !
4  3  0.00500 0  0  !  sp for O5-C12-O1  b-a = 4 deg
10  3  0.00500 0  0  !  sp for C12-O1-C19 b-a = 4 deg
11  3  0.00500 0  0  !  sp for O1-C19-C20 b-a = 4 deg
21  3  0.02812 0  0  !  sp for C16-C15-C14-C13 t-a = 22.4 deg
22  3  0.02812 0  0  !  sp for C15-C14-C13-C12 t-a = 22.4 deg
23  3  0.02812 0  0  !  sp for C15-C16-C12-O5  flap = 22.4 deg
24  3  0.02812 0  0  !  sp for C16-O5-C12-O1   t-a = 22.4 deg
25  3  0.02812 0  0  !  sp for C16-C15-C14-O3  t-a = 22.4 deg
26  3  0.02812 0  0  !  sp for C13-C14-C15-O4  t-a = 22.4 deg
27  3  0.02812 0  0  !  sp for C15-C14-C13-O2  t-a = 22.4 deg
28  4  0.02812 0  0  !  sp for C14-C15-C16-C17 t-a = 44.8 deg
29  4  0.02812 0  0  !  sp for C15-C16-C17-O6  t-a = 44.8 deg
30  7  0.02812 0  0  !  sp for C13-C12-O1-C19  t-a = 358 deg
31  7  0.02812 0  0  !  sp for C12-O1-C19-C20  t-a = 358 deg
32  7  0.02812 0  0  !  sp for O1-C19-C20-C21  t-a = 358 deg
33  3  0.02812 0  0  !  sp for C19-C20-C21-C22 t-a = 22.4 deg
34  3  0.02812 0  0  !  sp for C21-C22-C19-O8 flap = 22.4 deg
35  4  0.02812 0  0  !  sp for O8-C19-C20-O9   t-a = 44  deg
36  4  0.02812 0  0  !  sp for C19-C20-C21-O10 t-a = 44  deg
37  4  0.02812 0  0  !  sp for C19-O8-C22-C23  t-a = 44  deg
38  4  0.02812 0  0  !  sp for O8-C22-C23-O11  t-a = 44  deg
39  4  0.02812 0  0  !  sp for C21-C20-C19-C18 t-a = 44  deg
40  4  0.02812 0  0  !  sp for C20-C19-C18-O7  t-a = 44  deg
41  8  0.01400 0  0  !  sp for Rx rotation = 358  deg
42  8  0.01400 0  0  !  sp for Ry rotation = 358  deg
43  8  0.01400 0  0  !  sp for Rz rotation = 358  deg
44  5  0.17000 0  0  !  sp for Tx translation = 5.5 A
45  5  0.13600 0  0  *  !  sp for Tz translation = 4.4 A
0 3.0 10.00 0.25 0.80 0.20 0.10 0.70 0.005 0.0 0.0 !  powder data
SUC1.dat !  file contaning F2hkl
```

Table 2. Input data for sucrose test

13. Conclusions

The power of evolutionary algorithms in resolving difficult crystal structures from diffraction data has been discussed evidencing the convenience of basing the approach on internal coordinates. A specific procedure has been implemented, having the following main features: i) it is based on internal coordinates (this reduces considerably the number of variables and their uncertainty); ii) it uses discretized coordinates and binary structure encoding (metaheuristic approach); iii) the procedure is designed as a constraint satisfaction problem so incorporating the numerous a priori information available; the constraints are considerd both in the initial step of the procedure (formation of the initial population of solutions based on Montecarlo methods) and in the subsequent steps (breeding of the population).
The ideas above afforded a Fortran-language computer program suitable for any kind of molecular structure and available free of charge. The evolutionary procedure has been inserted into a general-purpose program, entirely based on internal coordinates. The program, presently tested on known structures (in difficult conditions), runs well. It has however the disadvantage of needing numerous parameters (presently assigned by the user) which could be assigned automatically in more evolute versions of the program.

14. References

Arnott, S. & Wonacott, A. J. (1966). *Polymers*, 7, 157-166

Back, T. (1996). *Evolutionary Algorithms in Theory and Practice: Evolution Strategy, Evolutionary Programming, Genetic Algorithms*, Oxford University Press, USA.

Bondi, A. J. (1964). *J. Phys. Chem.* 68, 441.

Cotton, A., Wilkinson G., Murillo, C.A., & Bochmann, M. (1999). *Advanced Inorganic Chemistry*, 6th Ed. J. Wiley & Sons, New York.

Davis, L (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold.

Eyring, H. (1932) *Phys. Rev. 39*, 746-748.

Feng, Z. J. & Dong, C. (2006) *J. Appl. Crystallogr.*, 39, 615-617.

Giacovazzo, C. (2002). *Fundamentals of Crystallography*, Oxford University Press, Oxford.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Berkeley.

Goldstein, H. (1980) *Classical Mechanics* 2nd Ed. Addison-Wesley, Berkeley.

Goto, H, & Osawa, E. (1989) *J. Am. Chem. Soc., 111*, 8950-8951.

Hanson, A. J., Cheung, E. J. & Harris, K. D. M. (2007) *J. Phys. Chem. B*, 111, 6349-6356.

Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (1998) *Acta Crystallogr.*, A54, 632-645.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems: an introductory analysis with applications to Biology, Control, and Artificial Intelligence*, MIT press, USA.

Hynes, R. C., Le Page J. (1991). *J. Appl. Cryst.* 24, 352-354.

Ionita, M., Breaban, N., & Croitoru C. (2010) *Evolutionary Computation in Constraint Satisfaction* in *New Achievements in Evolutionary Computing*, edt. P. Kovosec, Intech edition, Vukovar, Croatia.

Immirzi, A. (2007) (*a*). *Chem. Inf. Model*, 47, 2263-2265.

Immirzi, A. (2007) (*b*) *J. Appl. Cryst.*, 40, 1044-1049

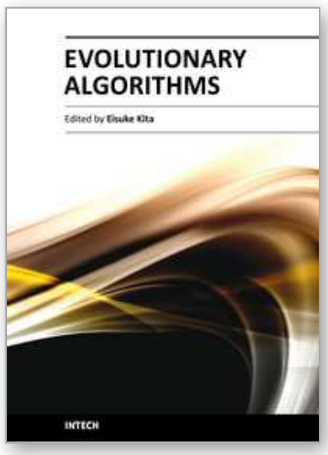
Immirzi, A., Erra, L., & Tedesco, C. (2008). *J. Appl. Cyrst.* 41, 784-790.

Immirzi, A., L. Erra, & Tedesco, C. (2009) *J. Appl. Cryst.*, 42, 810-814.

Lagrange, J. L. (1797). *Théorie des Fonctions Analytiques*, Imprimerie de la République, Paris.

Oganov, A. R., & Glass, C. W. (2006). *J. Chem. Phys.*, 124, art. 244704.
Okada, K., Noguchi, K., Okuyama, K., & Arnott, S. (2003). *Comput. Biol. Chem.*, 265-285
Polya, G. (1971). *How to solve it*. Princeton University Press.
Rietveld, H. M. (1967). *Acta Crystallogr.* 22, 151.
Rietveld, H. M. (1969). *J. Appl. Crystallogr.* 2, 65.
Sawicki, M. W., Li, N., & Gosh. D. (1999). *Acta Cryst. C*, C55, 1999, 425-427.
Smith, P. J. C. & Arnott, S. (1978). *Acta Cryst.*, A34, 3-11.
Stout, G. H., & Jensen, L. H. (1989). *X-ray structure determination, a practical guide*. J. Wiley & Sons.
Tadokoro, H. (1979). *Structure of Crystalline Polymers* J. Wiley & Sons, New York.
Weise, T. (2009). *Evolutionary Algorithms. Theory and Application*. URL : http://www.it-weise.de
Young, R. A. (1995) *The Rietveld Method*, IuCr Monograph on Crystallography, no. 5

IntechOpen



Evolutionary Algorithms

Edited by Prof. Eisuke Kita

ISBN 978-953-307-171-8
Hard cover, 584 pages
Publisher InTech
Published online 26, April, 2011
Published in print edition April, 2011

Evolutionary algorithms are successively applied to wide optimization problems in the engineering, marketing, operations research, and social science, such as include scheduling, genetics, material selection, structural design and so on. Apart from mathematical optimization problems, evolutionary algorithms have also been used as an experimental framework within biological evolution and natural selection in the field of artificial life.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Attilio Immirzi, Consiglia Tedesco and Loredana Erra (2011). Evolutionary Algorithms in Crystal Structure Analysis, Evolutionary Algorithms, Prof. Eisuke Kita (Ed.), ISBN: 978-953-307-171-8, InTech, Available from: http://www.intechopen.com/books/evolutionary-algorithms/evolutionary-algorithms-in-crystal-structure-analysis

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen