doi.org/10.1002/hlca.202200059

RESEARCH ARTICLE

HELVETICA chimica acta

Crystallography without Crystals: A Structural Study of Fakein

Angelo Gavezzotti*[a]

[a] Professor of Physical Chemistry (retired), University of Milano, Milano, Italy, e-mail: angelo.gavezzotti@unimi.it

Dedicated to the memory of *Jack D. Dunitz*

Software from the MiCMoS Molecular Simulation platform (sites.unimi.it/xtal_chem_group/) and from the *CryGen* set of programs (www.angelogavezzotti.it) is used to prepare a molecular structure from a line formula for a non-existent molecule, dubbed 'fakein'. Possible crystal structures are generated in chiral space groups $P2_1$ and $P2_12_12_1$, and for racemate in space groups P1− and $P2_1/c$. Lattice energies and crystal densities are in line with experimental values for organic molecules. Chiral crystals are less stable, and may not easily form hydrogen bonds, while bulk crystal stabilization is provided mostly by polarization-dispersion. Hydrogen bonding and crystal packing are discussed in terms of molecular pairing modes. Links to possible experiments and to absolute crystal structure prediction are outlined, but the main aim here is to demonstrate that relevant solid-state information can be derived if that goal is not achieved, actually even if the real substance is not available.

Keywords: chemoinformatics, computational chemistry, intermolecular potentials, MiCMoS platform, molecular simulation, organic crystals, structure prediction.

It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.

Niels Bohr, cited in A. Pais, The Genius of Science, *Oxford University Press, 2000, p. 24; as reported in J. D. Dunitz,* 'Gleanings', *private communication.*

Introduction

One of the many aspects of the astonishing evolution of X-ray crystallography in the last decades is the sharp decrease in the mass of necessary materials. In the 1970's, doing a single-crystal structure determination meant taking photographic plates that had to be decoded by pencil and paper transcripts, later punched on cards. Computer processing ended in hundreds of large printed-paper sheets; writing an article involved a paper manuscript, then typeset proofs, and eventually a nice bundle of journal reprints. Later on, diffraction intensities began to be loaded directly into a computer, with punched cards replaced by still pretty heavy magnetic tapes. By the mid 1980's, hard disks and word processors appeared, but there still was a lot of computer-printed paper around, and libraries bulged with tons of printed journals. With the advent of CCD technology, an X-ray diffraction pattern never emerged from its electronic limbo computer displays and bits and digits 'written' on a device nowadays the size of a few centimeters. When **electronic submissions of journal articles** began with the new century, paper disappeared altogether. The only solid trace of a structure determination is today a tiny batch of crystal samples. Remarkably, no step for the production of a published paper requires them to be shown.

The purpose of the present contribution is to explore ways in which even that last little piece of material evidence, an actual crystal, can be dispensed with, at least for organic substances. Computer simulation can reveal many facts about the relationships between a molecule and its possible crystal environment, even in absence of the tiny speck of matter that, after all, no one really cares to see. The

Supporting information for this article is available on the WWW under https://doi.org/10.1002/hlca.202200059

test is conducted using a nonexistent representative organic compound, aptly named 'fakein'. A molecular diagram is devised, a model is built from scratch by standard geometries, and crystal structures are calculated. To add a little spice to the presentation, a fictional account of plausible experimental sources of the test substance is included, and appropriate crystal sample selection and screening procedures are outlined.

With a little luck, the computer could turn out a reasonable approximation to the real crystal structure, cell dimensions, space group and atomic positions. But even without that, virtual crystallography provides reliable guesses at crystal polymorphism, perhaps even better than experiment, and also hints at solid state properties like packing modes and hydrogen bonding, crystal density, enthalpy of sublimation, and other downstream physical and thermodynamic properties.

Preparation of the Test Compound

Imaginary from Real World, Case 1

Fakein (*Figure 1*) was extracted from *Artemisia fictitia*, a bush that grows spontaneous in the tropical region of the State of Erewhon. A substantial amount of the dried plant was ground to powder and dissolved in a variety of solvents. The target compound was isolated by HPLC and further purified by repeated recrystallizations, affording a few grams of clean product (M.p. 380 K). Elemental analysis plus Mass Spectrometry confirmed a brute formula $C_{26}H_{21}Cl_2BrN_2O_9$. Given its natural origin, the compound was assumed to be optically pure. Further high-throughput recrystallization steps produced several batches of neat crystalline materials, including individuals with various shape and dimensions in excess of 0.2 mm, presumably suitable for single-crystal X-ray diffraction.

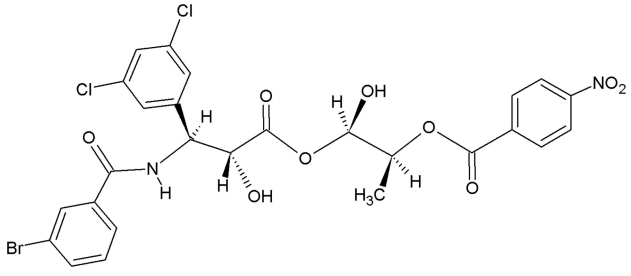

Figure 1. The fakein molecule. Ring positions and the main backbone vaguely resemble the non-cage part of taxol.

Imaginary from Real World, Case 2

Compound M31NGC224© (*Figure 1*), internal code name Fakein, was proposed as a potential lead into the development pipeline at the NEMOpharma Company. A synthetic route was planned with the help of an Artificial Intelligence software, reaching the target compound in good yield with a reasonable number of steps. As expected, however, the enantiomeric excess of the isomer shown in *Figure 1* was rather poor. In fact, extensive high-throughput recrystallization protocols yielded only a rather sparse crop of microcrystalline material, just a few milligrams, that under the microscope revealed the concomitant presence of acicular and tabular individuals. No separate melting could be observed, and the conglomerate mass melted at 387–390 K.

Modeling (Actual)

The molecular scheme for the virtual crystallography experiment is shown in *Figure 1*, taking broad inspiration from taxol, deleting the cage part, and adding a dash of substituents on the phenyl rings (Cl, Br, nitro). Module *Retcor* of the MiCMoS environment[1] (sites.unimi.it/xtal_chem_group/), a highly performing molecule builder, was used to produce Cartesian orthogonal coordinates for all the 61 atoms. Numerical values were input for the coordinates of the central C–C(=O)–O–C fragment, by simple sine and cosine calculations, while all other coordinates were generated through a bond distance-bond angle-torsion angle procedure or, for the benzene hydrogens and substituents, by having them lay on the bisector of the internal CCC angle. Geometric parameters were given average, standard values, taking care to introduce some realistic 'noise': for example, benzene C–C distances were between 1.387 and 1.393 Å; benzene angles between 118 and 122°; tetrahedral angles between 108 and 111°. The key problem was the assignment of torsion angles, for which the following assumptions were made: *1*) all benzene rings are planar; *2*) the central aliphatic or C–O–C chain is extended all *trans*; *3*) the bromobenzene ring is twisted by 30° out of the CONH amide plane, as in benzamide itself; *4*) the dichlorobenzene ring is constrained to avoid short contact between its *ortho*-hydrogens and the buttressing carbonyl oxygens; *5*) the nitrophenyl system is planar and coplanar with the COO group. Actual torsion angles had a ±2–3° spread around the 0, 60, 180 or 300° required by *cis* or *trans* geometries or by staggered aliphatic centers. All

assumptions so far are reasonable ones, stemming from personal experience after many decades of perusal and consideration of molecular structure and crystal packing modes in the Cambridge Structural Database. All manipulations were made easier by the program Schakal,[2] an unsurpassed molecular visualization tool that also allows instant spotting of critically short contacts, promptly corrected by refining torsion angles in a simultaneous use of Retcor.

The only critical problem was the placement of the two alcohol hydrogens, because the torsional barrier along the C–O(H) bond is nearly zero so that the O–H vector is free. One of the OH groups is oriented for a weak intramolecular interaction with a neighboring oxygen, while the other is made to point straight outwards of the main molecular frame, in a position hopefully conducive to intermolecular hydrogen bonding The heavily screened amide group is unavailable as a hydrogen bond acceptor or donor. The final result is shown in *Figure 2*: the two almost orthogonal phenyl rings are supposed to impart an awkward shape and to be a challenge to optimal packing arrangement.

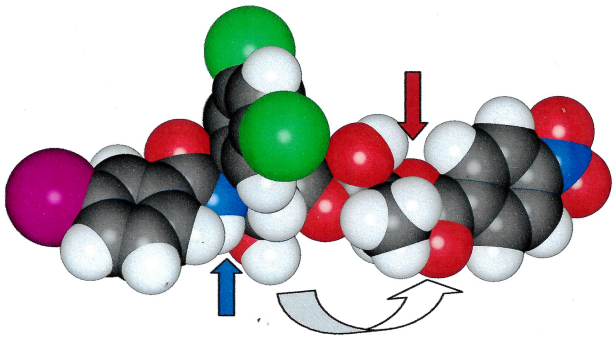

Figure 2. Molecular structure of fakein: oxygen red, nitrogen blue, chlorine green, bromine magenta. The red arrow (down) points to the site of a weak O–H···O intramolecular interaction. The blue arrow (up) points to the hindered NH group, unavailable for intermolecular hydrogen bonding. The curved arrow joins the donor O–H and the acceptor O=C sites oriented for intermolecular hydrogen bonding.

Fakein Crystallography

Optically Pure Compound of Natural Origin (Imaginary)

The material was passed on to the Department of Service Crystallography, where a preliminary screening of cell dimensions and space group for 30 specimens from different crystallization batches was carried out. This task (seldom if ever carried out in routine X-ray analysis) nowadays takes just a few minutes for each sample, but the reward is the immediate identification of polymorphs that may well escape visual observation – a different crystal form need not have a different crystal shape. In fact, two polymorphs were consistently detected, in space groups $P2_1$ and $P2_12_12_1$ (*Table 1*). Data collecting and data processing allowed a routine crystal structure determination for both polymorphs. A unique, fleeting specimen of dimensions <0.05 mm, picked from a batch kept under vacuum at low temperature, allowed only a few minutes of *Bragg* data collection, after which peaks broadened, due to an irreversible transition, presumably to an amorphous state. The new crystal structure ($P2_1$ HB, *Table 1*) could be determined by a set of intensities just sufficient to show the essentials of the packing motif. This turned out to be a 'disappearing' polymorph,[3] as the search for other samples of the same kind was discontinued after many frustrating efforts.

Crystals from Synthesis (Imaginary)

With top-notch equipment of the facility where NEMOpharma is outsourcing its X-ray diffraction operations, the smallness of the specimens did not prevent single-crystal analysis. After screening of 15 individual samples out of different recrystallization batches, two centrosymmetric polymorphs were detected in the obvious space groups $P1-$ and $P2_1/c$ (*Table 1*), showing that crystallization from the mother liquor out of the synthetic path could only occur by recognition between optical isomers of opposite handedness. The reasons of this selective process and

Table 1. Crystallographic data for the structures of fakein from imaginary experiments (actually, from simulations). The pseudo-refcode is in the style of CSD depositions (see *Supporting Information*). Lengths in Å, angles in degrees.

	Pseudo-refcode	*a*	*b*	*c*	α	β	γ	Density [g/cc^{-3}]	Packing coefficient
$P2_1$	FAKEIN01	16.09	6.964	15.27	–	60.7	–	1.46	0.66
$P2_12_12_1$	FAKEIN03	16.30	6.976	26.74	–	–	–	1.44	0.65
$P2_1$ (HB)	FAKEIN04	12.15	10.46	15.05	–	71.3	–	1.20	0.55
P1–	FAKEIN	6.995	13.47	18.68	95.8	55.0	90.9	1.52	0.69
$P2_1/c$	FAKEIN02	17.87	11.64	16.02	–	105.8	–	1.36	0.62

of the failure of observing crystal forms other than centrosymmetric ones, remain obscure, but the well-known attitude of the inversion center in promoting close packing, testified by the fact that it appears in about 70% of the crystal structure of organic compounds, is likely to have played a role. Data collecting and routine data processing for both polymorphs were successfully carried out.

Computational Crystallography (Actual)

Since the compound does not exist in reality, all crystal structures in *Table 1* were generated from the molecular model by the public domain *CryGen* set of programs.[4] The inherent procedure is a revamping of a very old idea,[5] briefly summarized as follows. The input consists of a rigid molecular model in Cartesian orthogonal coordinates, plus a few steering parameters, that control the required space group and the resolution of the search grids. The automaton then picks the 'leading' symmetry operator of each space group and produces a few energy-optimized clusters around it, dimers for inversion center and ribbons for a screw axis. In space groups with four equivalent positions, these primary clusters are further input to a search for the optimal action of a second operator, *e.g.*, in $P2_1/c$, a centrosymmetric dimer is re-optimized under the action of a screw axis to obtain a molecular layer. These clusters or layers are then expanded by projecting three vectors in space to yield a complete crystal structure in fractional atomic coordinates. These raw structures, whose number may run into thousands, are then input to extensive cycles of lattice energy minimization as a function of rigid-body degrees of freedom and of lattice parameters, interspersed with sorting out of duplicates. Each final structure carries the same information as from the output of an X-ray structure refinement, except for thermal parameters. The intermolecular potentials are the atom–atom CLP set, explicitly optimized for organic crystals,[6] while no intramolecular energy terms are needed since the molecule must be kept as a rigid body; this is the main shortcoming of the approach, so that in principle the whole process should be repeated for each plausible molecular conformation. The generated structures are 'zero-kelvin' or temperature-less ones, there being no account of lattice vibrations in the minimization procedure. Further detail and necessary algebra can be obtained from the *CryGen* manual[4] (free download of software and documentation).

As needed to mimic the imaginary experimental work, two chiral space groups for resolved material, $P2_1$ and $P2_12_12_1$, and two centrosymmetric ones for racemic material, P1− and $P2_1/c$, were explored by the *CryGen* robot. The final crop consisted of 23 structures in P1−, 14 in $P2_1$, 16 in $P2_1/c$, 6 in $P2_12_12_1$. *Figure 3* summarizes the results in the usual form of an energy/density plot. The structures with the most stabilizing lattice energy in each space group, plus a peculiar one for the 'disappearing' polymorph (*Table 2*), were chosen as corresponding to those described in the fictitious experiments (*Table 1*); they are available in *Supplementary Material* in the format of .cif files. Had the experiment been carried out, lattice energies of observed crystal structures could have been calculated in a few minutes from cell parameters and atomic positions, feeding the experimental .cif files into the

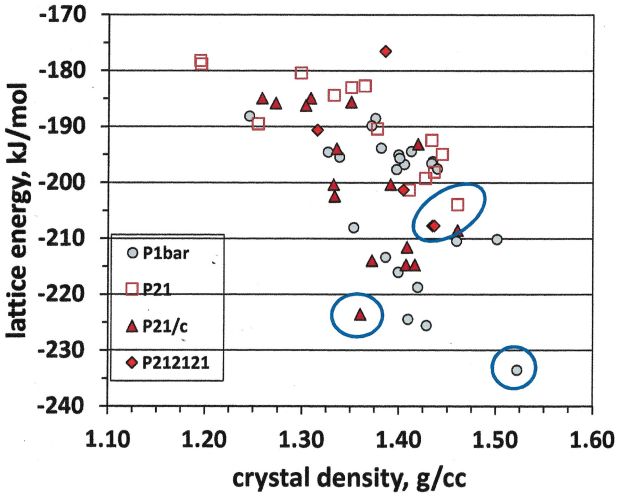

Figure 3. Results of the crystal structure generation runs for fakein. Circles denote the most stable polymorphs in each space group. One *P*1− structure stands out with most stable energy and highest density.

Table 2. Lattice energies (kJ mol^{-1}) for the crystal structures in *Table 1*: Coulombic, polarization, dispersion, repulsion, cell dipole energy in polar structures, and total lattice. Atom–atom CLP potentials.[6]

	*E*coul	*E*pol	*E*disp	*E*rep	*E*cell dipole	*E*(latt)	Hydrogen bond
resolved							
$P2_1$	−2	−60	−231	95	−7	−204	no
$P2_12_12_1$	−20	−60	−228	99	0	−208	no
$P2_1$ HB	−38	−47	−186	92	0	−179	yes
racemic							
P1−	−24	−65	−275	130	–	−234	yes
$P2_1/c$	−38	−54	−247	116	–	−224	yes

Retcif/Crysaa modules of the MiCMoS environment.[1] Incidentally, this cheap and easy procedure is recommended after all X-ray work, because the *Crysaa* module checks all intermolecular features with renormalized hydrogen atom coordinates and points out suspicious contacts, unlikely densities and packing coefficients, *etc.* Such checks are not included in the normal cif-checking procedures for deposition.

Further information on the packing characteristics of fakein comes from the plot (*Figure 4*) of the shortest cell axis as a function of lattice energy for the structures in *Figure 3*. For monoclinic structures, the shortest cell axis coincides almost always with the unique *b* axis, the direction of the screw operator. No fakein crystal may have a cell axis shorter than 7 Å, a feature than can appear in P1−, $P2_1$ and $P2_12_12_1$. In the popular space group $P2_1/c$, the minimum repeat distance is longer, 8 Å, but in the most stable structure, the shortest cell axis is even longer, 11.8 Å. *Figure 4* shows no correlation between short cell axes and stability, because strictly stabilizing interactions seldom occur by pure translation.

Discussion

Crystal Structure Prediction vs. Crystal Structure Generation

There is a large literature devoted to Computer Crystal Structure Prediction, CSP, performed by methods that are conceptually similar to *CryGen* but much more sophisticated in scope of application and procedures for the evaluation of energies.[7,8] Proper CSP includes a generation stage and a ranking and sorting stage, and also takes care of molecular flexibility, vastly increasing the number of structures generated. Its lattice energy evaluation is massively more demanding of computational resources; yet, CSP is only partially successful in matching experiments. For the present purpose, a much more restricted crystal structure generation is applied, in a study where no actual experiment is involved, and no comparison is possible. That many crystal structures with acceptable densities and high cohesive energy can be generated for *any* molecular substrate, of any composition, shape or size, is no novelty, having been verified on countless occasions.[7] Each point in *Figure 3* is a crystal structure whose lattice energy is a minimum in the restricted phase space of rigid-body motion and cell dimensions, and within the limits of reliability of the applied atom–atom potentials. The one outstandingly favorable crystal structure in *Figure 3* could be denoted as 'predicted' for the fakein molecule by the usual CSP recognition standards. Points within a 10% range in energy and density from the minimum energy one might qualify as possible polymorphs according to the observed statistics on differences among polymorph crystals.[9,10]

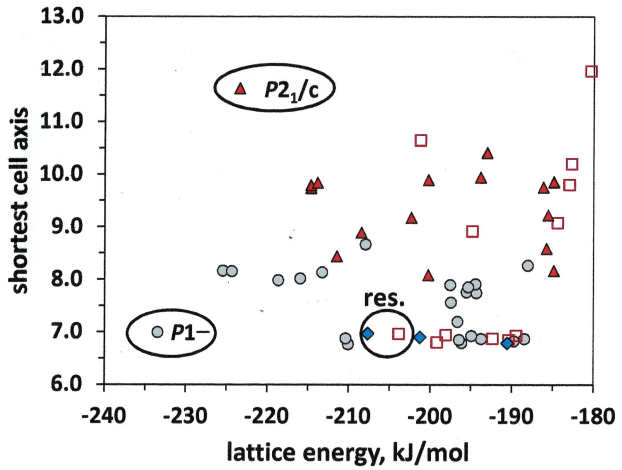

Figure 4. Shortest cell axis *vs.* lattice energy for the structures in *Figure 3*. No crystal structure can have a cell edge shorter than 7 Å, but otherwise note a total lack of correlation. Circles denote the best structures for the two centrosymmetric and the two chiral (resolved) space groups.

Structural Chemistry of the Fakein Solid State

Table 2 shows that the main bulk of the interaction energy in the fakein crystals is due to diffuse dispersion-polarization potentials. In fact, even hydrogen bonded structures are more stabilized by these forces, while the *Coulombic* part, the characteristic part of hydrogen bonding, is but a minor component. This situation is actually much more frequent than usually thought and especially in very large molecules, looking at hydrogen bonding as the main packing force can be misleading.[11] Of course, even more misleading could be looking after evanescent interactions such as C–H···(O,N,Cl) proximities.

Figure 5 shows a comparative crystal-packing tracker for the four stable crystal structures. Such graphs of pairwise molecule–molecule energies are a unique quantitative structural fingerprint.[12] The two centrosymmetric space groups are identical in the first determinant at 7.1 Å, but then differ considerably in the rest of the pairing energy spectrum. The two non-centrosymmetric structures are instead very similar, to the point that the orthorhombic form could almost be considered a structure modulation, with two cell

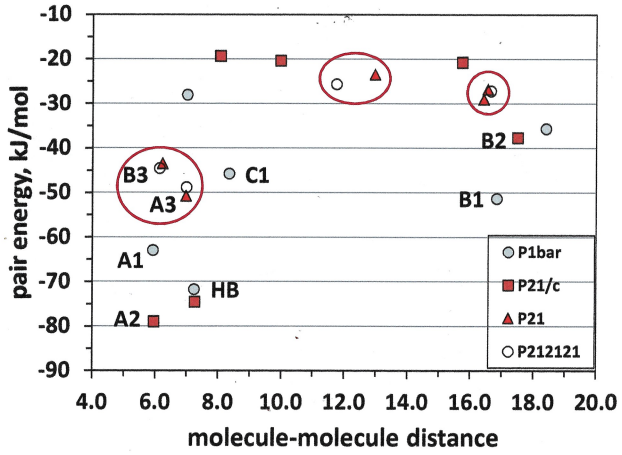

Figure 5. The packing energy spectrum of the four polymorphs. Each point denotes the interaction energy between a reference molecule and one of the surrounding molecules in the crystal (a structure determinant), at the given distance between centers of coordinates. Labels A, B, C, HB: see *Table 3* and *Figures 6*–9. The circles enclose the nearly overlapping determinants for the two chiral space groups.

parameters identical to those of the monoclinic form, and a near doubling of the third (*Table 1*). The data in *Figure 5* are shown numerically in *Table 3*, with detail of the kind of energy responsible for cohesion in each molecular pair.

Part of the reason for the enhanced stability of centrosymmetric forms is the formation of a doubly hydrogen bonded pair, possible only by total inversion (*Figure 6*). This motif is present, almost identical, in the best P1− and $P2_1/c$ generated structures, not surprisingly because it is the most obvious, if not the only possibility for hydrogen bond formation. Note how even in hydrogen bonded pairs the quantitative leader of cohesion is polarization–dispersion. Besides, in both space groups, a large part of stabilization is due to inversion operations that join molecules in terms of bulk, non-localized interactions (*Figure 7*). *Figure 8* shows the configuration by which the two almost orthogonal phenyl group are taken care of in the packing process; what may look like an awkward feature turns out in fact to be accommodated quite conveniently.

The superiority of P1− over $P2_1/c$ suggests that screw/glide symmetries are not so congenial to the fakein molecule. It is obviously impossible to form a double hydrogen bond as in *Figure 6* over a screw operator; indeed, the best $P2_1$ structure has no hydrogen bonds, and owes its main stabilization to a combination of screw- and translation-related couplings (*Figure 9*) with non-localized *Coulombic*-dispersion interactions (*Table 3*). Indeed, the screw operation leads to a net destabilizing *Coulombic* interaction, most likely an artifact of the approximate description of the atom-atom formulation that neglects penetration energies, that nevertheless sends a signal of the scarce influence of *Coulombic* terms. Reaching a configuration that includes single hydrogen bonds is a packing *tour de force* (*Figure 10*), found among the less stable ones by the computational robot. *Table 3* shows a strong OH···O=C hydrogen bond with high pairing energy, and a weak O···O=N interaction, but the

Table 3. Numerical detail with energy partitioning (kJ mol^{-1}) for the data shown in *Figure 5*. Label of the determinant: see *Figures 6*–9. 'Disappearing' polymorph: see *Figure 10*.

Distance [Å]	Symmetry	*Coulombic*	Polarization-dispersion	Total	Label
P1−					
7.23	invers.	−22	−98	−72	HB
5.92	invers.	−13	−71	−63	A1
16.83	invers.	−6	−83	−51	B1
8.33	invers.	−10	−95	−46	C1
$P2_1/c$					
5.96	invers.	−23	−103	−79	A2
7.25	invers.	−25	−99	−75	HB
17.50	glide	−3	−57	−38	B2
$P2_1$, $P2_12_12_1$					
6.96	transl.	−12	−62	−51	A3
6.22	screw	+6	−73	−43	B3
disappearing					
6.84	screw	−25	−86	−72	H… OC
12.95	screw	−12	−36	−34	H… ON

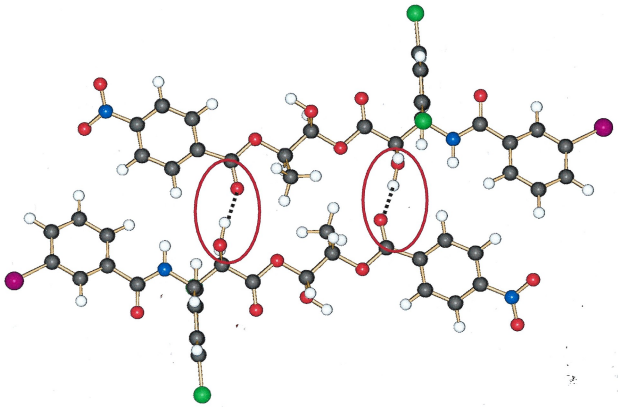

Figure 6. The doubly hydrogen bonded pairing over a center of symmetry in the best P1− and $P2_1/c$ crystal structures (determinants 'HB' in *Figure 5* and *Table 3*; symmetry 1−*x* ,−*y*, −*z* and −*x*, −*y*−1, −*z*, 1.89 and 1.82 Å, resp.).

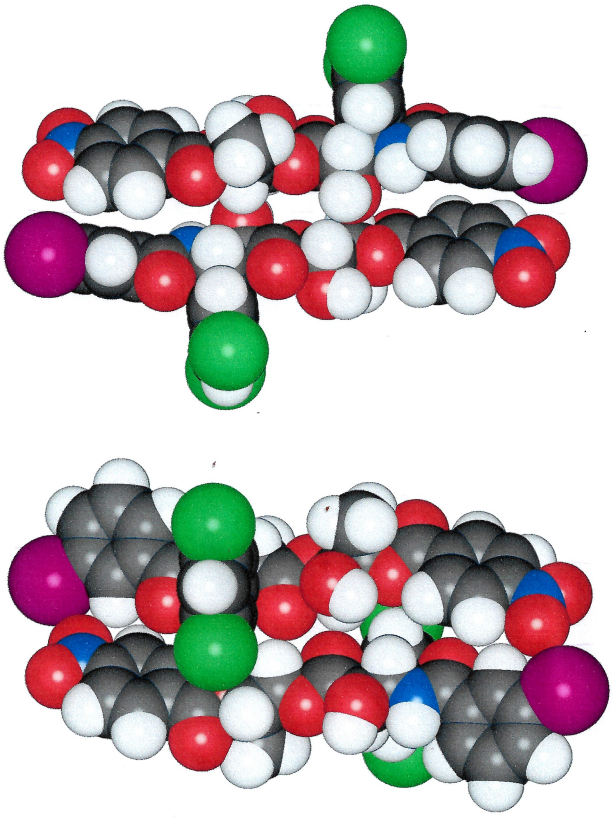

Figure 7. Examples of very stable centrosymmetric arrangements without hydrogen bonding: top, $P1-$, bottom, $P2_1/c$, symmetry $-x, -y, -z$. Such structures (determinants 'A1, A2' in *Figure 5* and *Table 3*) are stabilized by polarization and dispersion energies.

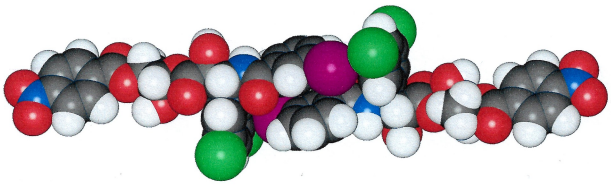

Figure 8. A centrosymmetric pair in the P1− crystal structure (determinant 'B1' in *Figure 5* and *Table 3*; symmetry $1-x, -y-1, -z-1$) that takes care of packing the two almost orthogonal phenyl rings, by a core-core stacking and two rim-core contacts.

awkward shape of the resulting couples precludes a further translational stabilization, as appears from the gaps in *Figure 10*. The result is a low lattice energy and density, as might be appropriate for a kinetically favored crystal form. This structure was described as a 'disappearing polymorph' in the virtual experimental section.

Crystallography without Crystals

The subset of theoretical chemistry based on non-quantistic simulation has grown to a stage where its results compete with, and sometimes improve upon experimental results, yielding valuable chemical information that proceeds from rigorous procedures based on sound physics. The applied intermolecular potentials are often calibrated on large sets of experimental properties of organic crystals. The whole setup, although totally empirical, can be trusted as a reliable expert system (or what might nowadays be called a piece of Artificial Intelligence).

In the present instance, the densities of crystal structures complying with the inescapable close-packing principle, that is with a packing coefficient of 0.6–0.7, are *ca.* 1.50 g cm^{-3}, can be trusted to be within a few percent points of the actual value (at the low temperature limit). Statistical studies on thousands of organic compounds[13] show that experimental sublimation enthalpies for 60-atom molecules range from 150 to 230 kJ mol^{-1}, the value increasing with molecular polarity and hydrogen bonding, so that fakein is likely to be at the upper limit of the distribution. Also, experimental sublimation enthalpies of alcohols with 40–70 atoms can be interpolated by $\Delta H(subl) = 3.4\ N(atoms)$, or 210 kJ mol^{-1} for fakein, perfectly in line with the values shown in *Table 2*. The sublimation enthalpy is in fact a reliable prediction, not only in absence of the actual substance, but also without need for time-consuming thermodynamic measurements, often affected by large uncertainties. The computational study shows that non-centrosymmetric crystal structures must be somewhat less stable than centrosymmetric ones, a valuable information when trying to crystallize a compound in different conditions of isomer resolution. The results also fall in the grey area of *Wallach*'s rule, according to which resolved crystals should be less dense than racemic crystals – a rule observed except when it is violated.[14,15] As a matter of fact, the best crystal structure is centrosymmetric and has the highest density and lattice energy, although density differences are at the margin of significance.

For fakein, the structural survey predicts that a real crystal is likely to have a 7 Å cell edge, and that no cell edge can be below that value. Coupling modes and hydrogen bonding possibilities are clearly borne out by the simulations, along with the predominance of dispersion energies over more localized or directional contributions. The results pertain to one conformation of the molecule under study, but aside from the fact

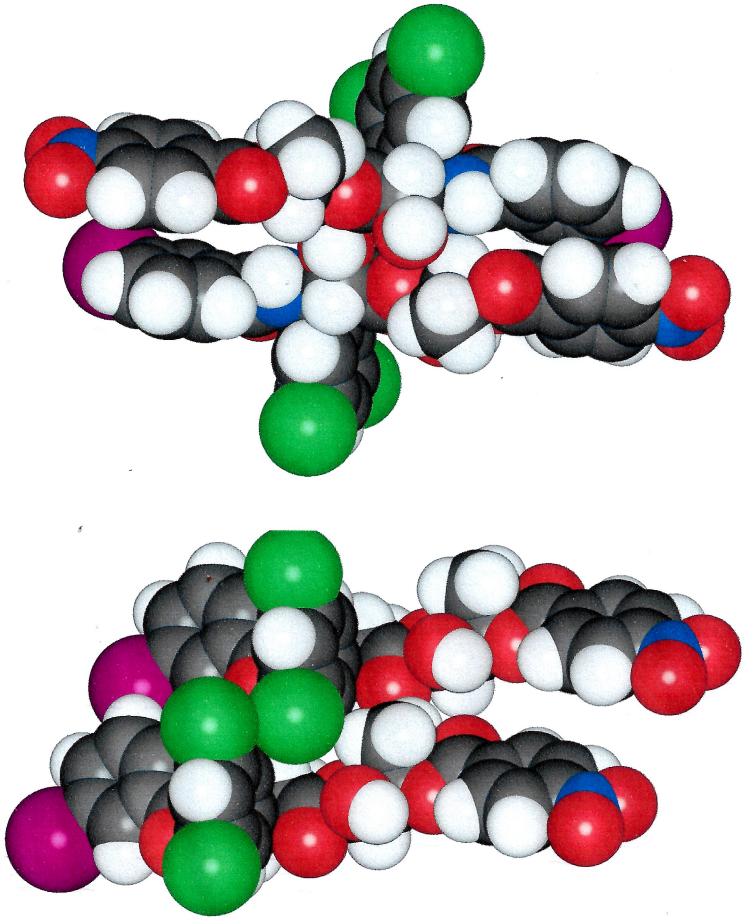

Figure 9. The main motifs in the $P2_1$ crystal structure: top, screw pair (determinant B3 in *Figure 5* and *Table 3*) seen approximately down the twofold axis; bottom, coupling by translation (determinant A3). The intermolecular gaps give a visual impression of the difficulty of adapting the molecule to a non-centrosymmetric environment.

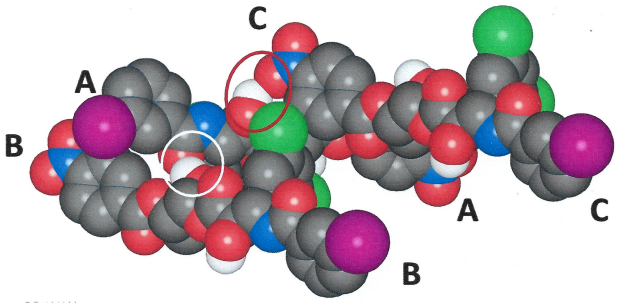

Figure 10. The $P2_1$ 'disappearing' polymorph: molecule A–A (partly hidden) accepts a O–H···O=C hydrogen bond from molecule B–B (1.78 Å, lower white oval) and donates a weak O–H···O=N hydrogen bond (2.03 Å, upper red oval) to molecule C–C, along a screw dyad. Only alcohol hydrogens are shown.

that other conformations for the present case are very unlikely, coupling some survey of the intramolecular energy landscape with repeated runs of the crystal structure generator would be well within reach of an even modest present-day computer platform. The resulting energy/structure landscape has intrinsic valuable information even if the holy Grail of exact crystal structure prediction is not attained. Having a plausible crystal structure, one can proceed to simulate surface properties and crystal morphologies,[16] important for stability and bioavailability, or to derive lattice vibrations and other thermodynamic, spectroscopic and heat conduction properties, important for materials science, by dedicated and extremely efficient downstream software.[17] These properties may change in different polymorphs, but any plausible structure can give an order-of-magnitude estimate and a preliminary idea of what one can expect from a given material.

Concluding Remarks

This paper describes a structural study of a nonexistent molecule from a scheme drawn on paper to its three-dimensional model and to its possible crystal structures. Hints are given as to how crystallography without crystals may yield crystal densities, enthalpies of sublimation, and the structures of possible polymorphs, along with downstream access to many of their solid-state properties. Remarkably, the whole job has been carried out using public domain software for molecular simulation and nothing more than an ordinary laptop computer, on which no production run exceeded a few hours' time. The whole process does not require specialized personnel, being within reach of organic chemists and crystallographers who may acquire the necessary expertise by following a limited set of instructions available in detailed manuals.

To add some *Dunitz* flavor, in the paradoxical spirit of his peer review of *Mendeleev*'s Periodic Table,[18] imaginary experimental routes to the same goal have been introduced with imaginary partners: *Artemisia fictitia* hardly needs comment, Erewhon (spell it backwards) is *Samuel Butler*'s non-existent country, M31NGC224, that mimics the labels usually given to pharmaceutical leads, is the catalog number of the Andromeda Galaxy, while *NEMOPharma* is a figment of the author's imagination, as could be guessed by its very name. Of course, the number and types of'experimental' polymorphs found were set *a posteriori* from the computational study.

Some might consider this work a spoof, or even worse, a potential scam. Plugging a computational structure from *CryGen*, or from any other available crystal structure predictor, into the experimental pipeline, requires just a few smart actions. Structure factors *F*(*calc*), calculated from simulated coordinates, could be matched to a set of 'observed' structure factors generated by some *ad hoc* algorithm, for example $F(obs) = (rand - 0.5) \cdot D_f \cdot F(calc)$, with $0 < rand < 1$ a random number and D_f a damping factor that can be adjusted to obtain any desired value of the crystallographic discrepancy index *R*. A bit more of make-up could come from a few cycles of '*F*(*obs*)'−*F*(*calc*) least squares, to generate pretended standard deviations, weighted *R*'s, anisotropic thermal factors, and the rest of the data that could fake a perfectly acceptable .cif file for deposition and publication. What prevents such malicious actions from being perpetrated is the very low value of their return, in times where a real structure determination takes minutes, against the high risk of forever losing a reputation if caught.

Are we heading toward a fully virtual research world? In parallel to the disappearance of solid matter in X-ray work, tangible objects are disappearing from the publication process. Authors are now required to do their own typesetting on a template supplied by the publisher, and Journals are in electronic format, so everything goes from a computer output to some other computer input/output. More and more often, and more and more explicitly, authors are asked to pay substantial money for having their papers published, through an astute trap called Open Access. OA grants well supported authors an enormous visibility advantage over their have-not colleagues in a Catch-22 loop, and also grants the Publisher a steady income without all the fuss about establishing and renewing subscriptions. Virtual crystallography dispenses with crystals; it looks like publishing is now dispensing with what once used to be called publication.

Supplementary Information

The *pseudo*-cif files for the crystal structures mentioned in *Table 1*.

Acknowledgements

The author is grateful to *Barbara* and *Jack Dunitz* for their very presence and for the inspiring atmosphere of so many evenings at their Küsnacht home. A twenty year cooperation with *Jack* resulting in 16 papers coauthored, plus three papers written for dedicated Journal issues and three lectures given at events in his honor, have been this author's salt of life.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

References

[1] A. Gavezzotti, L. Lo Presti, S. Rizzato, 'Molecular dynamics simulation of organic materials: structure, potentials and

the MiCMoS computer platform', *CrystEngComm* **2022**, *24*, 922–930.
[2] E. Keller, 'SCHAKAL92, A Program for the Graphic Representation of Molecular and Crystallographic Models', University of Freiburg, 1993.
[3] J. D. Dunitz, J. Bernstein, 'Disappearing Polymorphs', *Acc. Chem. Res.* **1995**, *28*,193–200.
[4] www.angelogavezzotti.it.
[5] A. Gavezzotti, 'Generation of Possible Crystal Structures from the Molecular Structure for Low-Polarity Organic Compounds', *J. Am. Chem. Soc.* **1991**, *113*, 4622–4629.
[6] A. Gavezzotti, 'Efficient computer modeling of organic materials. The atom–atom, Coulomb–London–Pauli (AA-CLP) model for intermolecular electrostatic-polarization, dispersion and repulsion energies', *New J. Chem.* **2011**, *35*, 1360–1368.
[7] A. Gavezzotti, 'Crystal structure prediction from molecular structure: Highlights and shadows', Chapter 5, in 'The Crystalline States of Organic Compounds', Ed. A. Gavezzotti, Elsevier, New York, 2021.
[8] A. J. Cruz-Cabeza, 'Crystal structure prediction: are we there yet?', *Acta Crystallogr. Sect. B* **2016**, *72*, 437–438.
[9] A. Gavezzotti, G. Filippini, 'Polymorphic Forms of Organic Crystals at Room Conditions: Thermodynamic and Structural Implications', *J. Am. Chem. Soc.* **1995**, *117*, 12299–12305.
[10] J. Bernstein, 'Polymorphism in Molecular Crystals', 2nd Edition, Oxford University Press, Oxford, 2020.
[11] J. D. Dunitz, A. Gavezzotti, 'Molecular Recognition in Organic Crystals: Directed Intermolecular Bonds or Non-localized Bonding?', *Angew. Chem. Int. Ed.* **2005**, *44*, 1766–1787.
[12] J. D. Dunitz, A. Gavezzotti, 'Toward a Quantitative Description of Crystal Packing in Terms of Molecular Pairs: Application to the Hexamorphic Crystal System 5-Methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile', *Cryst. Growth Des.* **2005**, *5*, 2180–2189.
[13] J. S. Chickos, A. Gavezzotti, 'Sublimation Enthalpies of Organic Compounds: a Very Large Database with a Match to Crystal Structure Determinations and a Comparison with Lattice Energies', *Cryst. Growth Des.* **2019**, *19*, 6566–6576.
[14] C. P. Brock, W. B. Schweizer, J. D. Dunitz, 'On the Validity of Wallach's Rule: On the Density and Stability of Racemic Crystals Compared with Their Chiral Counterparts', *J. Am. Chem. Soc.* **1991**, *113*, 9811–9820.
[15] A. Gavezzotti, S. Rizzato, 'Are Racemic Crystals Favored over Homochiral Crystals by Higher Stability or by Kinetics? Insights from Comparative Studies of Crystalline Stereoisomers', *J. Org. Chem.* **2014**, *79*, 4809–4816.
[16] M. Moret, A. Gavezzotti, 'The crystalline state of rubrene materials: intermolecular recognition, isomorphism, polymorphism, and periodic bond-chain analysis of morphologies', *New J. Chem.* **2022**, *46*, 7626–7637.
[17] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, B. Kirtman, 'Quantum-mechanical condensed matter simulations with CRYSTAL', *WIREs Comput. Mol. Sci.* **2018**, *8*, e1360.
[18] J. I. Seeman, 'Peer Review of *Mendeleev*'s 1869 Breakthrough Paper: 'I Suggest Eliminating the Table…', *Helv. Chim. Acta* **2019**, *102*, e1800177.

Received May 30, 2022
Accepted June 28, 2022

