NMR Crystallography: Toward Chemical Shift-Driven Crystal Structure Determination of the β-Lactam Antibiotic Amoxicillin Trihydrate

DATASET · APRIL 2013

CITATIONS
3

READS
21

3 AUTHORS, INCLUDING:

Joao Rocha
University of Aveiro
455 PUBLICATIONS **9,819** CITATIONS
SEE PROFILE

Luís Mafra
University of Aveiro
86 PUBLICATIONS **1,254** CITATIONS
SEE PROFILE

CRYSTAL GROWTH & DESIGN

Article

pubs.acs.org/crystal

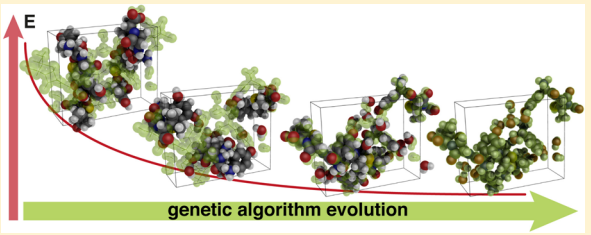
NMR Crystallography: Toward Chemical Shift-Driven Crystal Structure Determination of the β-Lactam Antibiotic Amoxicillin Trihydrate

Sérgio M. Santos,* João Rocha,* and Luís Mafra*

Department of Chemistry, CICECO, University of Aveiro, 3810-193 Aveiro, Portugal

Supporting Information

ABSTRACT: We report a new strategy for NMR crystallography of multiple-component molecular crystals in which ^{1}H NMR chemical shifts enter directly in the structure generation step, governed by a genetic algorithm. Chemical shifts are also used in the structure-refinement step as pseudoforces acting on the models, leading to the lowest-energy structure. This methodology, which avoids the use of time-consuming *ab initio* chemical shift calculations, is successfully applied to powdered amoxicillin trihydrate, a widely used β-lactamic antibiotic.



INTRODUCTION

The recent developments in *ab initio* structure determination have led to the emergence of the so-called SMARTER crystallography (structure elucidation by combining magnetic resonance, computational modeling, and diffractions[1]) as a way to derive the three-dimensional crystal packing of powdered samples. This approach benefits from the high sensitivity of NMR to the local chemical environments of molecules in crystals. In particular, ^{1}H chemical shifts are a powerful source of structural information for studying hydrogen-bonds and the orientations of small molecules.[2] Recently, the structure of a rigid molecule (thymol) was solved from a combination of experimental and calculated NMR parameters,[3a] in conjunction with crystal structure prediction (CSP) methods.[3b] In the approach presented here, we introduce ^{1}H isotropic chemical shifts directly into the CSP step, based on a genetic algorithm (GA) approach, which also includes additional information derived from powder X-ray diffraction. This provides a fast tool for crystal structure determination, based on classical force-fields and semiempirical chemical shift approaches, avoiding the time-consuming high-level density functional theory (DFT) calculations. For demonstrating the concept, the multiple component crystal amoxicillin trihydrate β-lactam antibiotic is used (Figure 1).

METHODS

The progress in the prediction of crystal packing of organic and inorganic systems using CSP approaches was recently reviewed by Day et al.[4] Genetic algorithms (GA) are a class of heuristic algorithms commonly used in CSP, which mimic the natural evolution of living systems.[5] In GA, the candidate solutions to a problem (e.g., optimum crystal packing) are encoded in a population of chromosomes, each characterized by a value determining its fitness within the population. As the GA evolves, certain individuals (parents) are selected according to their fitness and combined to produce new candidate individuals (offsprings). The fittest offsprings are reinserted in the main population, replacing the least fit parents. As in natural systems, an offspring may undergo mutation of some of its genes, yielding a genotype containing information not inherited from its parents (maintaining gene diversity in the population). At the end of the GA evolution, the fittest individual presumably encodes the best (global) solution to the problem. In the context of CSP, the use of *a priori* known experimental or theoretical information will speed up the evolution of the population toward the solution, since it considerably narrows the search-space of the procedure.

So far, NMR crystallography of organic molecular crystals has essentially followed a two-step strategy: structure proposal and structure validation. In the former, numerous structures are generated fulfilling certain experimental restraints, namely, space group information and unit cell parameters. In the second step, the generated structures are ranked and validated against experimental data, often using time-consuming calculations of NMR parameters, based on high-level DFT. Often, the experimental NMR data play a passive role in the entire crystal-structure determination process, solely acting as reference (target). The methodology proposed here embeds the NMR data directly into the structure-generation step, using GA for global minimization, relying on a three-step strategy: (i) chemical-shift constrained CSP (structure generation); (ii) simulated annealing of the CSP-derived structures (structure refinement), and (iii) tight-binding DFT (DFTB) geometry optimization of the annealed structures (structure minimization). In (i), cell parameters and space group information, forthcoming from powder X-ray diffraction, are combined with experimental ^{1}H chemical shifts in a genetic algorithm, generating a set of possible crystal packing structures. With this information known beforehand, the entire cell is constructed upon application of the adequate space group symmetry operations to the packed asymmetric unit. In (ii), the generated structures are refined through molecular dynamics simulated annealing (SA), governed by forces derived from a classical force field with an additional term. This

Received: January 15, 2013
Revised: March 23, 2013
Published: April 4, 2013

dx.doi.org/10.1021/cg4002785 | *Cryst. Growth Des.* 2013, 13, 2390−2395

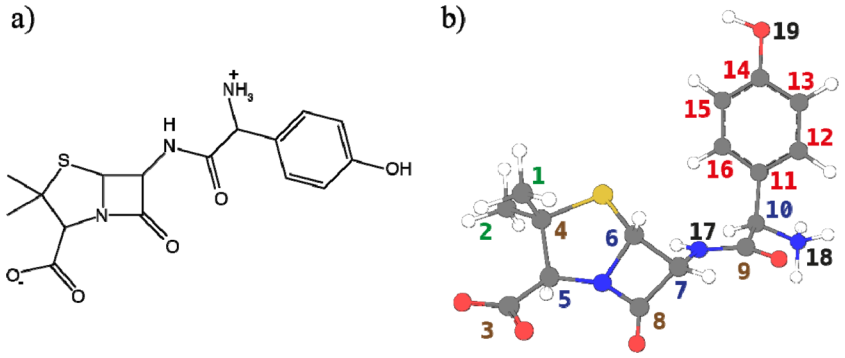


Figure 1. (a) Structural formula and (b) atom labeling of amoxicillin trihydrate (zwitterionic form).

term accounts for the deviation of the measured and on-the-fly calculated chemical shifts, yielding a new set of structures fulfilling the experimental restraints. In (iii) the structures are energy-minimized by DFTB. This procedure is demonstrated on the determination of the crystal-structure of powdered amoxicillin trihydrate, $C_{16}H_{19}N_3O_5S{\cdot}3(H_2O)$, which crystallizes in space-group $P2_12_12_1$ with four distinct molecules in the unit cell ($Z = 4$) (CSD code AMOXCT10).[6]

In GA, step (i), the configuration of the asymmetric unit is completely described by a chromosome encoding the molecular conformation, center-of-mass position, and orientation of the amoxicillin molecule and the three water molecules in the unit cell. In GA terminology, the translation of the chromosome yields the asymmetric unit, which upon application of the adequate symmetry operations, generates the entire unit cell. Hence, three genes per molecule are needed for encoding the three Euler angles characterizing the molecular orientation, three genes per molecule for the coordinates of the molecular position, and four additional genes encoding the variable dihedral angles of amoxicillin (Figure S15 in Supporting Information). The problem is, thus, completely described by a chromosome with 28 genes. Real-valued chromosomes were chosen, each gene x having a value $0 \leq x < 1$, representing a fractional coordinate of the molecule's center-of-mass or an Euler/dihedral angle (recovered upon multiplication by π or 2π).

GA searches the global minimum of the potential energy surface by finding the raw fitness (E_{total}) of each generated trial structure. E_{total} (eq 1) is defined as the sum of the lattice energy ($E_{lattice}$) of the system with a term accounting for the deviation between the experimental (δ_{exp}) and the on-the-fly calculated (δ_{calc}) ^{1}H chemical shifts (E_δ):

$$E_{total} = E_{lattice} + E_\delta \tag{1}$$

$$E_\delta = \sum_h \begin{cases} k(\delta^h_{calc} - \delta^h_{exp} + \varepsilon_\delta)^2, & \delta^h_{calc} \leq \delta^h_{exp} - \varepsilon_\delta \\ 0, & |\delta^h_{calc} - \delta^h_{exp}| < \varepsilon_\delta \\ k(\delta^h_{calc} - \delta^h_{exp} + \varepsilon_\delta)^2, & \delta^h_{calc} \geq \delta^h_{exp} + \varepsilon_\delta \end{cases} \tag{2}$$

$E_{lattice}$ being described by the classical biomolecular force-field GAFF,[7] which comprises intramolecular (bonds, angles and dihedrals) and intermolecular (Coulomb and van der Waals) components, the latter having contributions from all unit cells (see Supporting Information for additional details). In E_δ, the sum runs over all the protons h for which experimental chemical shifts (δ_{exp}) are available. The parameter ε_δ controls the width of the flat-well potential and defines a threshold for the accepted error in the δ_{calc} values. The on-the-fly calculation of δ_{calc} requires a fast methodology, amenable to integration within a classical molecular mechanics/dynamics code. Inspired by previous protein solution NMR work,[8] the calculated chemical shift of a given proton is parametrically partitioned into

$$\delta_{calc} = \delta_{bond} + \delta_{ef} + \delta_m + \delta_{rc} \tag{3}$$

where δ_{ef}, δ_m, and δ_{rc} are the through-space electric-field (ef), group magnetic anisotropy (m), and aromatic ring current (rc) contributions, respectively, and δ_{bond} is the through-bond (sum of electronic dia- and paramagnetic) contributions to the proton chemical shift. The through-bond contribution is assumed to depend only on the molecular topology (i.e., atomic connectivities) and, thus, independent of the molecular conformation and crystal packing. δ_{bond} is assumed to be proportional to the partial atomic charge of the probe proton and held constant throughout the course of the procedure.[9] The electric-field contribution to the chemical shift (δ_{ef}) of a given proton bonded to atom X (X−H) is described by the Buckingham model.[10] The group-magnetic anisotropy contribution (δ_m) is described by the McConnell model[11] (Figures S7−S10 in Supporting Information) and originates from the asymmetric electron distribution in the functional group, which, under an external magnetic field, generates an internal electric current inducing a local magnetic field opposing the external one. The position of the probe nucleus relatively to the functional group causes its electronic (de)shielding. Finally, and like δ_m, the application of an external magnetic field induces an electron current in the aromatic rings, giving rise to an induced local magnetic field opposing the external one. To describe the contribution (δ_{rc}) of these currents to the proton chemical shift, the Haigh-Mallion model[12] was adopted (see Figures S11−S12 in Supporting Information).

The isotropic chemical shifts are only calculated for the asymmetric unit protons because all symmetry-related nuclei have the same chemical shift. Only the chemical shift terms (eq 3) of the first neighbor shell (the 26 surrounding cells) are considered because the inclusion of additional shells is negligible. The chosen definition of E_{total} allows the introduction of experimentally derived constraints into the process of structure generation through the E_δ term,[13] perpetuating individuals that best fit the experimental data and which have evolved during the GA course. E_δ does not ensure, however, the physical and chemical meaning of the generated structures, as it only accelerates the evolution of the population toward the best global potential-energy minimum. The physicochemical meaning of the generated structures is ensured by the $E_{lattice}$ term, which avoids atom overlap and bond lengths, angles, and dihedrals incompatible with the molecular topology of the intervenient species. Although the water protons are not included in the calculation of E_δ, as their individual experimental chemical shift experimental cannot be measured separately, they do contribute to the amoxicillin chemical shifts via the δ_{ef} term. The $E_{lattice}$ contribution is fundamental for determining the correct packing of the water molecules, strengthening the applicability of classical biomolecular force-fields to CSP of molecular solids.

To further refine the candidate structures, a molecular dynamics SA, step (ii), is employed. If GA is allowed to proceed for a sufficiently long time, with adaptable parameters, the population converges to the final optimum crystal structure. This strategy requires finely tuned parameters and suitable simulation times to allow for global convergence. Hence, the SA step is used to let the systems overcome small potential-energy barriers, since the high-energy barriers are dealt with in step (i). The initial packing integrity found in GA is maintained during the SA step. Simultaneously, the E_δ component of the force-field drives the system toward a configuration obeying the imposed restraints.

dx.doi.org/10.1021/cg4002785 | Cryst. Growth Des. 2013, 13, 2390−2395

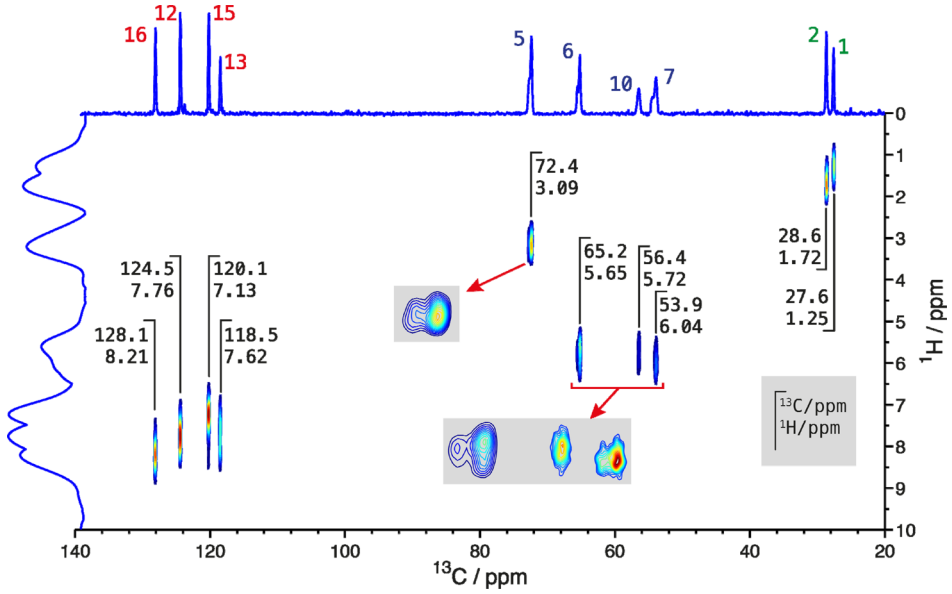


Figure 2. $^{1}H-^{13}C$ PRESTO-HETCOR NMR spectrum recorded at 400 MHz using a MAS rate of 10 kHz and ^{1}H CRAMPS decoupling along the indirect dimension (see Figures S2–S4 in Supporting Information); peak labeling according to Figure 1.

In step (ii), the force $\mathbf{F}_i$ acting on the ith particle of the system is given by

$$\mathbf{F}_i = -\frac{\partial E_{\text{total}}}{\partial \mathbf{r}_i} = -\frac{\partial E_{\text{lattice}}}{\partial \mathbf{r}_i} - \frac{\partial E_{\delta}}{\partial \mathbf{r}_i} \quad (4)$$

where $\mathbf{r}_i$ is the position vector of i. The equations of motion are integrated using a leapfrog scheme, with a time step of 1 fs to correctly account for the fastest X–H vibrational modes. The temperature of the system is controlled through direct scaling of the atom velocities. The structures are initially heated to 1000 K and cooled down to 0 K at a rate of 0.2 K/ps. The initial heating does not disturb significantly the initial packing found by GA because the solid-state nature of the system does not allow high mobility to the larger components (amoxicillin molecules), while largely influencing the arrangement of the water molecules. During SA, symmetry operations are also applied to each step of the coordinate update, resulting in a partially constrained molecular dynamics simulation.

At the end of the SA refinement step, the candidate solutions are geometry-optimized at the self-consistent-charge DFTB level (SCC-DFTB) implemented in the DFTB+ program,[14] step (iii). This approach allows for fast geometry optimization, providing a very good description of organic molecular systems, while requiring moderate computational resources. In our case, each geometry optimization takes ca. 5 min in a 2.6 GHz 8-core desktop computer (notably less than an equivalent pseudopotential planewave DFT optimization would take in the same computer).

RESULTS AND DISCUSSION

The ^{13}C CPMAS NMR spectrum of amoxicillin trihydrate (Figure S2 in Supporting Information) displays 16 resonances, thus confirming the presence of a single antibiotic molecule in the asymmetric unit. The two-dimensional $^{1}H-^{13}C$ PRESTO-HETCOR NMR spectrum (Figure 2) allows the assignment of the 14 through-bond $^{1}H\cdots^{13}C$ connectivities. The PRESTO block is based on an RN dipolar-recoupling symmetry-based sequence, which affords a selectivity similar to that of J-mediated experiments albeit with higher sensitivity. Full-resonance assignment is obtained by combining experimental and gauge included projector augmented wave (GIPAW[15]) calculated ^{1}H and ^{13}C chemical shifts (see Figure S5 and Table S1 in Supporting Information). The chemical shifts of the protons attached to carbons are used as reference (δ_{exp} in eq 2) during the GA. The chemical shifts of the remaining (−NH, $-NH_3$, and −OH) protons are not used because unambiguous resonance assignment is not possible.

The GA structure generation step uses a population of 100 individuals, with all genes initiated to uniformly distributed random numbers x in the interval $[0, 1[$. In each generation, 20 parents are chosen by the tournament selection method, with each tournament round consisting of five randomly selected individuals from the main population. Each parent couple is combined by crossover (homologous swap of genes) at one point, to yield two new offsprings. Each offspring undergoes mutation at a single random locus with probability $p_m \in [0.1, 0.02]$, the mutation being the sum of the old gene with a random number withdrawn from a Gaussian distribution with zero mean and width $m_w \in [0.5, 0.1]$. The newly formed offsprings are reinserted into the main population, in accordance with the elitist+fitness criterium: the 10 worst individuals of the main population are replaced by the 10 best offsprings. GA is allowed to evolve for 1000 generations, with p_m and m_w linearly scaled in the given intervals at each generation. Parameters for the calculation of E_{lattice} are taken from GAFF force field, with RESP[16] fitted partial atomic charges calculated at the HF/6-31G* level on a previously optimized amoxicillin molecule. The water molecules are described by the TIP3P model.[17] Each generated individual is energy minimized through 100 steps of steepest-descent optimization prior to the calculation of E_{total}. Details concerning the calculation of E_{δ} may be found as Supporting Information, alongside the description of the parametrization strategy. The threshold parameter ε_{δ}, in eq 2, is taken as 0.25 ppm, corresponding to a flat-well potential of 0.5 ppm width, and the force constant k is set to 1000 kcal mol^{-1} ppm^{-2} (this value is comparable to the force constants found in the GAFF force field).

Figure 3 shows the evolution of the fitness (E_{total}) of the fittest individual during 1000 generations over 5 distinct runs, with and without inclusion in the fitness function of the NMR chemical shift restraints, selected such that the final structures

dx.doi.org/10.1021/cg4002785 | Cryst. Growth Des. 2013, 13, 2390–2395

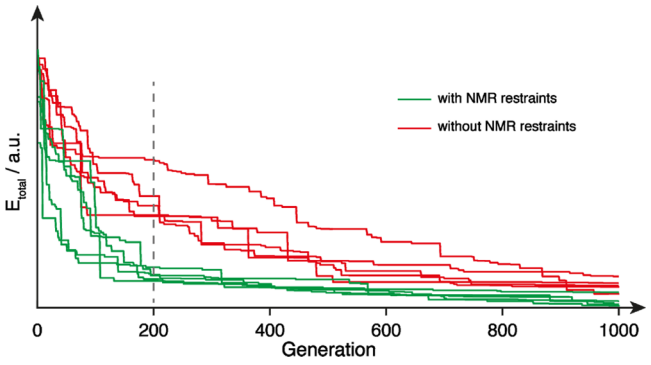


Figure 3. Evolution of E_{total} of the fittest individual during the course of the genetic algorithm over five distinct runs, with and without inclusion of chemical shift restraints (green and red curves, respectively). Apart from the inclusion of the chemical shift restraints, all running parameters are the same in the two sets (see also Figure S16 in Supporting Information).

are comparable to the reference X-ray structure and the final fitnesses (E_{total}) are similar for all runs. The inclusion of experimental data (E_δ) accelerates GA convergence toward the minimum energy, so that much fewer (hundreds) generations are required to obtain a reasonable set of molecular packings. In the shown data set, the runs with NMR restraints converge to the global energy minimum basin in ca. 200 generations, whereas runs without NMR restraints require significantly more generations. According to eq 1, one might expect runs with inclusion of NMR restraints to have E_{total} fitness values higher than homologous runs without inclusion of restraints, since $E_\delta \geq 0$. However, the E_δ penalty is compensated by lower $E_{lattice}$ terms because the chemical shift restraints drive the system toward configurations with a more favorable packing (lower $E_{lattice}$) quicker than when NMR restraints are absent, since individuals with high E_δ penalties are swiftly removed from the breeding pool. The E_δ term introduces an additional set of variables complementary to those found in the classical force field used in the calculation of the lattice energy. In the limit of all chemical shift-derived restraints being fulfilled ($E_\delta = 0$), the GA and SA evolution is entirely governed by $E_{lattice}$ of each individual, with the quality of the generated structures fully determined by the quality of $E_{lattice}$. One must note that when the global $E_{lattice}$ minimum basin is located (i.e., the ideal packing is found), the parameter ε_δ, that controls the width of the flat-well potential, makes E_δ tend to zero because, within the basin, the changes in the true chemical shifts (up to 0.5 ppm since we chose $\varepsilon_\delta = 0.25$ ppm) fall within the error of our semiempirical approach. Nevertheless, the tandem use of $E_{lattice}$ and E_δ allows for the swift generation of candidate packings because the chosen functional form of these terms and their gradients are easily evaluated by any modest desktop computer. One must note that convergence of the algorithm would also be observed without inclusion of NMR restraints, although at the cost of more iterations. The degeneracy of the $E_{lattice}$ is greater than that of $E_{lattice} + E_\delta$, and the sole use of $E_{lattice}$ as the figure of merit leads to the momentary trapping of the algorithm in local minima; more iterations are then needed to escape from them. Hence, the E_δ term lifts part of the lattice energy degeneracy, reducing the chances of the algorithm being trapped in local minima.

The 100 SCC-DFTB geometry-optimized lowest-energy drug structures obtained from five independent GA+SA+DFTB runs are very similar (Figure 4; atomic coordinates available as Supporting Information), with only minor

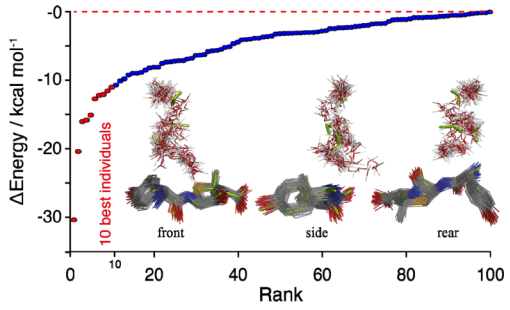


Figure 4. Plot of the ranked energies (relatively to the least fit structure) of the 100 SCC-DFTB geometry optimized lowest-energy structures, obtained from five independent GA+SA+DFTB runs (see Supporting Information for the coordinates of the 10 best structures). The inset shows the overlap of the 100 asymmetric units (water and drug molecules in the top and bottom rows, respectively), the asymmetric unit of the reference X-ray structure being depicted in green sticks.

variations in the backbone dihedrals (the major difference is observed in the position of the water molecules around amoxicillin). Thus, the plot of the ranked energies of these 100 structures seems to be determined by the location of the water molecules in the unit cell. To some extent, this reflects the mobility of the water molecules (the drug molecules are relatively rather immobile). Figure 5 illustrates the lowest/energy predicted structure overlapped with the single-crystal X-ray structure. The all-atom root-mean-square deviation (RMSD) between both structures is 0.608 Å (or 0.576 Å, if only the asymmetric unit is considered), which may be considered an excellent agreement, given that the entire procedure is based on low level, non *ab initio*, parametrized models.

At this point, some important considerations must be addressed. First, the entire procedure is based on a semi-empirical calculation of proton chemical shifts using parameters fine-tuned to correctly describe the GIPAW calculated counterparts; this latter calculation required knowledge of amoxicillin trihydrate's crystal structure which, in a real life application of the proposed methodology, would be unknown! This is by no means a problem, since the only requirement is that the semiempirical strategy accurately calculates chemical shifts. In such a case, models and parameters should also be transferable between systems, in a way similar to what is currently done in protein structure determination, where parameters depend solely on the nature of the amino acids and not on the protein itself (see, for instance, ref 8); nevertheless, and to ensure that this method is generally applicable, further work is required to develop and demonstrate models and parameters for semiempirical calculations of ^{1}H chemical shifts that are successful for a wide range of organic molecular crystals. Naturally, one can go fully *ab initio* and calculate chemical shifts from first principles, provided enough computational power is available, in which case this problem no longer exists (although the procedure can no longer be run in a single workstation due to the large number of structure evaluations, as was done in the present case). Second, we used GIPAW chemical shifts to aid in the assignment process, using the known crystal structure. In the real case of an unknown

dx.doi.org/10.1021/cg4002785 | *Cryst. Growth Des.* 2013, 13, 2390−2395

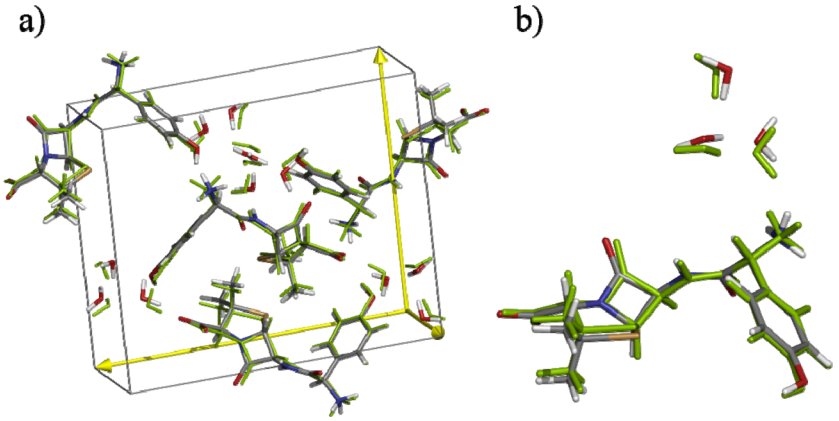


Figure 5. Predicted best-crystal packing of amoxicillin trihydrate: (a) entire unit cell, (b) asymmetric unit. The reference X-ray structure is shown as green sticks. RMSD between both structures is 0.608 Å (or 0.576 Å if only the asymmetric unit is considered).

crystal structure, the assignment step is a problem, since one can no longer map the calculated chemicals shifts to the measured counterparts. When possible, the assignment task can be facilitated using other NMR techniques that provide atomic connectivities, such as *J*-mediated NMR experiments if sensitivity and spin−lattice relaxation times are favorable (e.g 2D $^{13}C-^{13}C$ INADEQUATE). When such methods are unpractical, the assignment problem may be circumvented in the following way: take all measured chemical shifts and sort them (either ascendingly or descendingly); predict a new crystal packing and calculate the corresponding chemical shifts; take these and sort them in the same order as the measured ones; the sum (S^2) of the squared differences between matching values should provide a measure of the fit between the proposed and the real crystal structures (this will only work if one assumes that there are no two distinct configurations that yield the same sorted chemical shifts); as $S^2 \rightarrow 0$, one gains access to both resonance assignment and structure. We are currently exploiting this approach to the complete assignment of crystals with yet unknown structures. Finally, data derived from powder X-ray diffraction were used to constrain the unit cell and space group. Since such sort of information is not always readily available from diffraction techniques, the entire procedure can be easily modified to incorporate these new variables, as is already done in CSP techniques.

CONCLUSION

In conclusion, a new approach to SMARTER crystallography, which uses ^{1}H NMR chemical shifts in the crystal-structure generation step, was introduced and demonstrated with amoxicillin trihydrate powder. The procedure yields a set of optimized structures closely matching the single-crystal X-ray structure (RMSD ∼ 0.608 Å). This is a step forward toward the prediction of crystal structures, based on semiempirical NMR chemical shifts, without using time-consuming *ab initio* chemical shift calculations.

ASSOCIATED CONTENT

Supporting Information

Description of empirical chemical shifts models and parametrization procedure, coordinates of the final 10 lowest energy structures, 1D and 2D high-resolution NMR spectra, GIPAW chemical shift calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: (L.M.) lmafra@ua.pt; E-mail: (S.M.S.) sergio.santos@ua.pt; E-mail: (J.A.) rocha@ua.pt.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financed by FEDER, via "Programa Operacional Factores de Competitividade − COMPETE", and by FCT − "Fundação para a Ciência e a Tecnologia", PEst-C/CTM/LA0011/2011. We also thank FCT for their financial support to the Project PTDC/QUIQUI/100998/2008 and for postdoc grant SFRH/BPD/64752/2009 (S.M.S.).

REFERENCES

(1) (a) Martineau, C.; Cadiau, A.; Bouchevreau, B.; Senker, J.; Taulelle, F.; Adil, K. *Dalton Trans.* **2012**, *41*, 6232. (b) Brouwer, D. J. *Magn. Reson.* **2008**, *194*, 136.

(2) (a) Brown, S. P. *Solid State Nucl. Mag.* **2012**, *41*, 1. (b) Hansen, M.; Graf, R.; Sekharan, S.; Sebastiani, D. *J. Am. Chem. Soc.* **2009**, *131*, 5251. (c) Mafra, L.; Santos, S. M.; Siegel, R.; Alves, I.; Almeida Paz, F. A.; Dudenko, D.; Spiess, H. W. *J. Am. Chem. Soc.* **2012**, *134*, 71. (d) Dudenko, D.; Kiersnowski, A.; Shu, Jie.; Pisula, W.; Sebastiani, D.; Spiess, H. W.; Hansen, M. R. *Angew. Chem., Int. Ed.* **2012**, *51*, 11068.

(3) (a) Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2610. (b) Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *J. Am. Chem. Soc.* **2010**, *132*, 2564. See also Harris, R. K. *Solid State Sci.* **2004**, *6*, 1025 for further examples of the use of chemical shifts in structure determination.

(4) (a) Day, G. M. *Acta Crystallogr. B* **2009**, *65*, 107. (b) Day, G. M. *Crystallogr. Rev.* **2011**, *17*, 3.

(5) Reeves, C. R.; Rowe, J. E. *Genetic Algorithms: Principles and Perspectives - a Guide to GA Theory*; Kluwer Academic Publishers: Massachusetts, USA, 2003.

(6) Boles, M. O.; Given, R. J.; Gane, P. A. C. *Acta Crystallogr. B* **1978**, *34*, 461.

(7) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.

(8) See, for instance: (a) Hunter, C. A.; Packer, M. J.; Zonta, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 27. (b) Cross, K. J.; Wright, P. E. *J. Magn. Reson.* **1985**, *64*, 220. (c) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436.

(9) Wolff, R.; Radeglia, R. In *Encyclopedia of NMR*; Grant, D. M., Harris, R. K., Eds.; Wiley: Chichester, U.K., 1996; Vol. 2, pp 488−497.

(10) (a) Buckingham, A. D. *Can. J. Chem.* **1960**, *38*, 300. (b) Grayson, M.; Raynes, W. T. *Magn. Reson. Chem.* **1995**, *33*, 138.
(11) (a) McConnell, H. M. *J. Chem. Phys.* **1957**, *27*, 226. (b) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436.
(12) (a) Haigh, C. W.; Mallion, R. B. *J. Chem. Phys.* **1982**, *76*, 4063. (b) Christensen, A. S.; Sauer, S. P. A.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2078.
(13) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson. B* **1995**, *107*, 293.
(14) (a) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260. (b) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678. (c) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185.
(15) Pickard, C. J.; Mauri, F. *Phys. Rev. B* **2001**, *63*, 245101.
(16) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.