

Crystal structure prediction: are we there yet?

Aurora J. Cruz-Cabeza*

The University of Manchester, Sackville Street, Manchester M13 9PL, England. *Correspondence e-mail: aurora.cruzcabeza@manchester.ac.uk

Received 11 July 2016

Accepted 12 July 2016

Keywords: crystal structure prediction; polymorphism; lattice energies; Cambridge Structural Database.

Predicting the way a molecule will crystallize under a given set of conditions is a challenging task. Sketch out a molecule into a computer and run complex algorithms in the hope that the computational result matches that of the experiment. Being able to predict crystal structures computationally in an accurate and time-efficient manner could have enormous implications in industries that develop organic crystalline materials (e.g. pharmaceuticals). Crystal structure prediction (CSP) is, therefore, an ultimate dream in the context of pharmaceutical material sciences, but are we there yet?

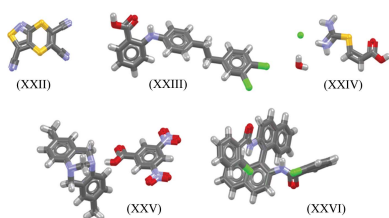
The development of CSP methods has been partly catalyzed by the blind test competitions organized at the Cambridge Crystallographic Data Centre. Experimental outcomes of crystallizations for a set of targets are held in confidence whilst molecular sketches are given to participants who are then allowed to run their computations over the course of nine months and submit their computer-generated crystal structures. The blind test was first run in 1999 (Lommerse *et al.*, 2000) and since then it has been regularly organized every 3 years (Motherwell *et al.*, 2002; Day *et al.*, 2005, 2009; Bardwell *et al.*, 2011). The results of the sixth blind test of CSP methods are being published in the present issue of *Acta Crystallographica Section B* (Reilly *et al.*, 2016). So, what have we learnt?

The CSP problem can be broken into two parts: (i) a sampling problem and (ii) a ranking problem. The sampling problem relates to the number of ways a given molecule can be packed in a three-dimensional space in the form of an ordered crystal. The ranking problem relates to the identification of the structure or structures, which will be observed experimentally from all those possible packings under a given set of conditions. Over the years, blind test targets have increased in complexity; but, what does complexity mean in this context? Sampling becomes more difficult as we increase the flexibility of the target molecule, as we increase the number of symmetrically independent molecules in the crystal and as we increase the number of components. A complex sampling, however, does not actually mean a complex ranking too. In fact, quantifying complexity at the ranking stage, *a priori*, is virtually impossible.

In the sixth blind test, five different targets were attempted (Fig. 1): a small rigid molecule (XXII), two significantly large and flexible molecules (XXIII and XXVI), a two-component cocrystal (XXV) and a hydrate of a salt (XXIV). In addition, target (XXIII) had the difficulty of crystallizing in five different packings (polymorphs), two of which each had two symmetry-independent molecules.

As in previous blind tests, the accuracy of methods employed and group experience played an important role in the success of the groups. The know-how is an important element as well as the quality of the sampling and ranking methods. Participants trying to implement newer methods often only attempted the simplest target (XXII), whilst only groups with more history and experience attempted all targets.

Consolidated sampling algorithms were able to generate the observed packing in all targets except for the salt hydrate, which was only generated by one group, and the structure of one of the polymorphs of target (XXIII), which was not generated by any method. The sampling complexity of this blind test has been enormous and tremendous progress has been achieved, but it has been highlighted (once again) that we still have some way to go. Hydrates of salts are rather common in pharmaceuticals as well as complex polymorphs. Given a molecule, the sampling of all possible solid forms is virtually infinite once we start accounting for polymorphs, cocrystals, salts, hydrates of salts, solvates of cocrystals... Whilst consolidated algorithms are now able to sample the packing of some of these forms (many of which are incredibly complex), we are still far



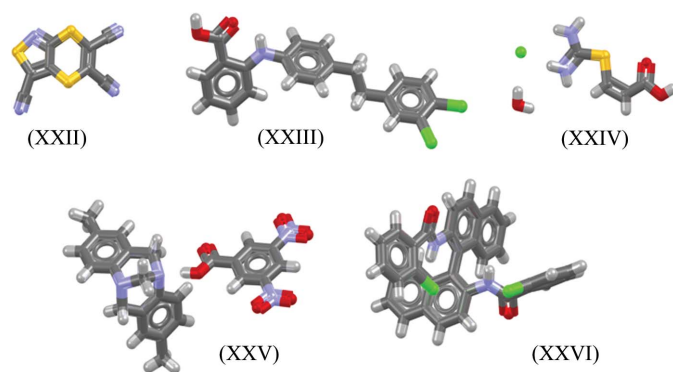


Figure 1
Targets in the sixth blind for crystal structure prediction.

from being able to do efficient computational screenings of the entire solid form universe of a given compound. We are at the point, however, at which these methods can complement and add value to experimental screenings.

With respect to rankings, most ranking methods are solely based on energy (lower-energy structures are therefore more likely to be observed experimentally). It was well established in previous blind tests that classic force-field models perform poorly at the ranking stage, so most groups have now adopted more sophisticated energy models including some based on density functional theory. Most targets would have been correctly ranked by various of the more sophisticated ranking methods, some of which have been intensively developed in the past years. This is a very impressive outcome, which is consistent with the findings in the fourth and fifth blind tests. The most stable polymorphs for target (XXIII) (*A* and *D*), albeit only identified under unusual conditions, were not correctly ranked by any method. In that regard, whilst there has been a tremendous advancement in ranking methods, further studies with complex polymorphic systems are a must for future editions of the blind tests and for future benchmarking studies. As discussed above, it is not yet clear when a particular target is going to present an especially difficult ranking stage. Why are most targets perfectly ranked with advanced energy methods but a few targets still misbehave?

Finally, we are still far from relating experimental conditions of crystallization to the obtained crystal structures. In fact, none of the blind test participants made use of the crystallization conditions for their predictions. In an ideal scenario, we would like to be able to predict structures computationally and then derive an experimental procedure to produce them in the laboratory. For this to occur, however, there is a considerable amount of fundamental work still needed to advance our understanding of nucleation and our ability to simulate it computationally. Whilst we are coming close to predicting the stable structure of a given compound, we still cannot answer why some compounds are polymorphic or which of the predicted forms can be realized experimentally and how. As the methods have been advancing, the actual purpose of the blind test has been evolving too from prediction of structure to prediction of solid-form landscapes and polymorphism.

The sixth blind test, more than ever before, represents an outstanding community effort with 25 groups, 92 authors and 52 institutions taking part. Beside that, extensive experimental form screenings were organized for two of the targets. Wow! This is a must read article!

References

- Bardwell, D. A., Adjiman, C. S., Arnautova, Y. A., Bartashevich, E., Boerrigter, S. X. M., Braun, D. E., Cruz-Cabeza, A. J., Day, G. M., Della Valle, R. G., Desiraju, G. R., van Eijck, B. P., Facelli, J. C., Ferraro, M. B., Grillo, D., Habgood, M., Hofmann, D. W. M., Hofmann, F., Jose, K. V. J., Karamertzanis, P. G., Kazantsev, A. V., Kendrick, J., Kuleshova, L. N., Leusen, F. J. J., Maleev, A. V., Misquitta, A. J., Mohamed, S., Needs, R. J., Neumann, M. A., Nikylov, D., Orendt, A. M., Pal, R., Pantelides, C. C., Pickard, C. J., Price, L. S., Price, S. L., Scheraga, H. A., van de Streek, J., Thakur, T. S., Tiwari, S., Venuti, E. & Zhitkov, I. K. (2011). *Acta Cryst.* **B67**, 535–551.
- Day, G. M., Cooper, T. G., Cruz-Cabeza, A. J., Hejczyk, K. E., Ammon, H. L., Boerrigter, S. X. M., Tan, J. S., Della Valle, R. G., Venuti, E., Jose, J., Gadre, S. R., Desiraju, G. R., Thakur, T. S., van Eijck, B. P., Facelli, J. C., Bazterra, V. E., Ferraro, M. B., Hofmann, D. W. M., Neumann, M. A., Leusen, F. J. J., Kendrick, J., Price, S. L., Misquitta, A. J., Karamertzanis, P. G., Welch, G. W. A., Scheraga, H. A., Arnautova, Y. A., Schmidt, M. U., van de Streek, J., Wolf, A. K. & Schweizer, B. (2009). *Acta Cryst.* **B65**, 107–125.
- Day, G. M., Motherwell, W. D. S., Ammon, H. L., Boerrigter, S. X. M., Della Valle, R. G., Venuti, E., Dzyabchenko, A., Dunitz, J. D., Schweizer, B., van Eijck, B. P., Erk, P., Facelli, J. C., Bazterra, V. E., Ferraro, M. B., Hofmann, D. W. M., Leusen, F. J. J., Liang, C., Pantelides, C. C., Karamertzanis, P. G., Price, S. L., Lewis, T. C., Nowell, H., Torrisi, A., Scheraga, H. A., Arnautova, Y. A., Schmidt, M. U. & Verwer, P. (2005). *Acta Cryst.* **B61**, 511–527.
- Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Cryst.* **B56**, 697–714.
- Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Lommerse, J. P. M., Mooij, W. T. M., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2002). *Acta Cryst.* **B58**, 647–661.
- Reilly, A. M., Cooper, R. I., Adjiman, C. S., Bhattacharya, S., Boese, A. D., Brandenburg, J. G., Bygrave, P. J., Bylisma, R., Campbell, J. E., Car, R., Case, D. H., Chadha, R., Cole, J. C., Cosburn, K., Cuppen, H. M., Curtis, F., Day, G. M., DiStasio Jr, R. A., Dzyabchenko, A., van Eijck, B. P., Elking, D. M., van den Ende, J. A., Facelli, J. C., Ferraro, M. B., Fusti-Molnar, L., Gatsiou, C.-A., Gee, T. S., de Gelder, R., Ghiringhelli, L. M., Goto, H., Grimme, S., Guo, R., Hofmann, D. W. M., Hoja, J., Hylton, R. K., Iuzzolino, L., Jankiewicz, W., de Jong, D. T., Kendrick, J., de Klerk, N. J. J., Ko, H.-Y., Kuleshova, L. N., Li, X., Lohani, S., Leusen, F. J. J., Lund, A. M., Lv, J., Ma, Y., Marom, N., Masunov, A. E., McCabe, P., McMahon, D. P., Meeke, H., Metz, M. P., Misquitta, A. J., Mohamed, S., Monserrat, B., Needs, R. J., Neumann, M. A., Nyman, J., Obata, S., Oberhofer, H., Oganov, A. R., Orendt, A. M., Pagola, G. I., Pantelides, C. C., Pickard, C. J., Podeszwa, R. I., Price, L. S., Price, S. L., Pulido, A., Read, M. G., Reuter, K., Schneider, E., Schober, C., Shields, G. P., Singh, P., Sugden, I. J., Szalewicz, K., Taylor, C. R., Tkatchenko, A., Tuckerman, M. E., Vacarro, F., Vasileiadis, M., Vázquez-Mayagoitia, A., Vogt, L., Wang, Y., Watson, R. E., de Wijs, G. A., Yang, J., Zhu, Q. & Groom, C. R. (2016). *Acta Cryst.* **B72**, 439–459.