MS46-P5 Reduction by deduction – How to prioritize crystallization experiments

Max Pillong[1], Corinne Marx[1], Wicker Jerome[2], Cooper Richard[2], Wagner Trixie[1]

1. Novartis Institute of Biomedical Research
2. Department of Chemistry, University of Oxford

email: max.pillong@novartis.com

Computational tools have found an increasing use throughout various topics and fields in chemistry. From the prediction of physicochemical properties[1], reaction types and products[2] or biological affinities towards macromolecules[3] up to the high-dimensional problem of crystal structure prediction, so-called *Machine Learning* models experience a steady increase in application. *Machine Learning* models are considered a collection of algorithms that can be used to find regularities, irregularities and correlations in data sets independent of their dimensionality. The resulting mathematical constructs can then be used to predictively characterize previously unseen data. In crystallography, recent studies have set out to use such algorithms in order to predict general crystallinity of small molecules[4], their crystallization propensity in different solvents[5], or the crystallization conditions of proteins[6].

The most important prerequisite for a high-performing machine learning model is a sufficiently large data set of high integrity. Unfortunately, due to the lack of negative results in commonly used collections like the CSD, crystallization data suited for machine learning tasks is sparse. Therefore, several years back we set out to record both positive and negative results for every crystallization attempt made in the analytics group at the Novartis Institute for Biomedical Research. This data is stored in an SQL data base equipped with a touch-screen based graphical user interface, enabling easy access for both experimenter as well as programmer.

Next to numerous statistical analyses, we set out to use machine learning for the prediction of suitable conditions to crystallize small, drug-like molecules. In particular, we focused on the prediction of organic solvents to crystallize a given compound in. While the individual performances of the first generation of machine learning models for this were rather frugal, a newly devised ensemble approach embodying additional data allows us to rationalize our crystallization experiments and thus significantly reduce the experimental effort required to yield crystalline materials.

[1] Alzghuol A., *et al.* (2014) *J Chem Inf Mod*el **54**, 3396-403

[2] Schneider N., *et al.* (2016) *J Med Chem* (10.1021/acs.jmedchem.6b00153)

[3] Reker D., *et al.* (2014) *PNAS* **111**, 4067-72

[4] Wicker J.P., Cooper R. (2015) *CrystEngComm* **17**, 1927-34

[5] Hosokawa K., *et al.* (2005) *ChemPharmBull* **10**, 1296-99

[6] Rupp B., Wang J. (2004) *Methods* **34,** 390-407

Keywords: crystallization, machine learning, database, statistical analysis

MS46-P6 Crystal structure, Hirshfeld surfaces and DFT computation of (E)-(5-methylfuran-2-yl) (morpholino) methanone oxime

Ersin Temel[1], Aydın Demircan[2], Medine Çolak[2]

1. Department of Electric and Energy, Technical Sciences Vocational High School, Giresun University, TR-28000, Giresun/Turkey
2. Department of Chemistry, Faculty of Arts and Science, Niğde University, TR-51240, Niğde, Turkey

email: ersin.temel@giresun.edu.tr

N-hydroxy amidoximes are one of the most important amidine derivatives with synthetic utility and various biological applications. They have been used extensively as starting materials for the preparation of nitrogen-rich heterocyclic compounds. Characteristically, they can also react or cyclize with electrophiles such as aldehydes, ketones, carboxylates, and acids. In practice, they have applications in drugs, dyes, polymers, and many other materials as well.

With this work, we provide a synthesis of *N*-hydroxy amidoksim from the reaction between furfurylimidoyl chloride and *N*-morpholine [1]. The structure of (Z)-(5-methylfuran-2-yl) (morpholino) methanone oxime was investigated with experimental (X-ray single crystal technique, NMR and FT-IR spectroscopic techniques) and theoretical (DFT) techniques. The compound crystallizes in monoclinic space group P21/c. Crystal structure is stabilized by inter-molecular O-H…N and C-H…O hydrogen bonds. The Hirshfeld surface was drawn for visualizing the van der Waals distances and to determine the interaction sites. It is understood that O-H..N and C-H…O type hydrogen bonds are dominant interactions on the packing. When 2D fingerprint plot are partitioned, H-H interactions are seen to be the most dominant interactions with percentage of 60.2. And then O-H/H-O interactions with 20.0% and C-H/H-C interactions with 12.0% come. The puckering amplitude for the six-membered ring was determined as Q=0.572(3)Å. The gas phase geometry optimization and vibrational frequencies calculations were carried out using density functional theory (DFT) incorporated in B3LYP with 6-311++(d,p) basis set. The detailed vibrational assignments were performed on the basis of the potential energy distributions (PED) of the vibrational modes. Additionally, HOMO-LUMO energy gap, natural bond orbital (NBO) analysis and nonlinear optical (NLO) properties of the compound were performed.

Acknowledgement: We would like to thank TUBITAK (PN: 113T136) and Giresun University for financial support of this work.

[1] A. R. Katritzky, L. Huang, M. Chahar, R. Sakhuja, C. D. Hall, *Chem. Rev.* **2012**, 112, 1633–1649; P. Vitale, S. Tacconelli, M. G. Perrone, P. Malerba, L. Simone, A. Scilimati, A. Lavecchia, M. Dovizio, E. Marcantoni, A. Bruno, P. Patrignani *J. Med. Chem.* **2013**, 56, 4277−4299; V. Mercalli, M. Giustiniano, E. Del Grosso, M. Varese, H. Cassese, A. Massarotti, E. Novellino, and Gian Cesare Tron, *ACS Comb. Sci.* **2014**, 16, 602−605.