CL1

New Table-Top Diffractometer Bruker D6 Phaser

NOVÝ STOLNÍ DIFRAKTOMETR BRUKER D6 PHASER

Boris Míč

Měřící Technika Morava

Nový stolní rentgenový difraktometr Bruker D6 PHASER představuje průlomové řešení pro rentgenovou difrakční analýzu v kompaktním provedení. Díky svým pokročilým funkcím a volitelným doplňkům nabízí výkon, který je běžně dostupný pouze u větších, laboratorních difraktometrů. Uživatelé mohou využít široké škály aplikací, od základní fázové analýzy po pokročilé studium struktury materiálů. D6 PHASER podporuje vysokou flexibilitu při konfiguraci experimentů a je ideální pro výzkum i průmyslové aplikace. Mezi jeho volitelné doplňky patří například automatický měnič vzorků, nízkoteplotní komory, a různé typy detektorů, které umožňují měření při specifických podmínkách. Navíc, díky své stolní velikosti, je D6 PHASER vhodný i pro menší laboratoře, kde je prostor omezený. Navzdory kompaktnímu designu nedochází ke kompromisům v přesnosti a citlivosti měření. Tento difraktometr je tedy ideální volbou pro uživatele, kteří potřebují výkonné a všestranné zařízení v dostupném a praktickém balení. D6 PHASER přináší vysokou úroveň inovace a spolehlivosti do světa rentgenové difrakce.

Studentská přehlídka I

SL1

FROM UNCERTAINTY TO MOLECULAR MECHANISM: MISSENSE MUTATIONS IN *BRAF* AND *MAP2K1* IN COGNITIVE DISORDERS

P. Havlickova, A. Koutska, I. Kuta Smatanova, M. Fenckova

Faculty of Science, University of South Bohemia in Ceske Budejovice, Branisovska 1760, Ceske Budejovice, 37005, Czech Republic

fenckm00@prf.jcu.cz

The RAS/MAPK signalling pathway is one of the most extensively studied pathways, mainly due to its fundamental role in the regulation of cell cycle, proliferation, or senescence. Mutations in RAS/MAPK are primary drivers of cancer [1]. They also contribute to developmental syndromes known as RASopathies. These syndromes are associated with body malformations of different severity and with impaired cognitive function. Individuals with these syndromes often experience intellectual disability (ID) and autism spectrum disorder (ASD) [2, 3]. In this study, we investigate *de novo* recurrent single-point missense mutations in the *BRAF* and *MAP2K1* genes (encoding kinases BRAF and MAP2K1), found in individuals with ID and ASD. They are currently classified as variants of unknown significance (VUS) [4]. It is crucial to investigate whether and how they impact the protein function, as some variants in RASopathies are known to increase while decreasing the kinase activity.

The variants were selected from large sequencing studies [4] using an initial dataset of all missense variants found in ID/ASD individuals. We considered the number of affected individuals with each variant and the presence of secondary mutations in other genes. Further, the variants were cross-validated for their association with ID or ASD with ClinVar Miner (https://clinvarminer.genetics.utah.edu/), SysNDD (https://sysndd.dbmr.unibe.ch/) and SFARI database (https://gene.sfari.org/).

We cloned the wild-type BRAF (BRAFwt), wild-type MAP2K1 (MAP2K1wt), and the corresponding VUS-containing variants. We express and purify the proteins for assessment of kinase activity and for investigating the effect of VUS on 3D structure with X-ray crystallography. This study aims to elucidate the pathogenicity of the novel mutations and the molecular mechanisms by which they lead to ID/ASD. This will help to improve diagnostics and find tailored treatments targets.

1. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. Cell, 100(1), 57–70. https://doi.org/10.1016/s0092-8674(00)81683-9.
2. Rauen K. A. (2013). The RASopathies. Annual review of genomics and human genetics, 14, 355–369. https://doi.org/10.1146/annurev-genom-091212-153523.
3. Geoffray, M. M., Falissard, B., Green, J., Kerr, B., Evans, D. G., Huson, S., Burkitt-Wright, E., & Garg, S. (2021). Autism Spectrum Disorder Symptom Profile Across the RASopathies. Frontiers in psychiatry, 11, 585700. https://doi.org/10.3389/fpsyt.2020.585700.
4. Wang, T., Kim, C. N., Bakken, T. E., Gillentine, M. A.,.Henning, B., Mao, Y., Gilissen, C., SPARK Consortium, Nowakowski, T. J., & Eichler, E. E. (2022). Inte-

grated gene analyses of de novo variants from 46,612 trios with autism and developmental disorders. Proceedings of the National Academy of Sciences of the United States of America, 119(46), e2203491119. https://doi.org/10.1073/pnas.2203491119

This work is supported by by a grant from the Czech Science Foundation (grant no. 23-07810S) and an EMBO Installation grant (grant no. IG-5310-2023) to M. Fenckova.

SL2

INVESTIGATING THE EFFECT OF NEURODEVELOPMENTAL DISORDER MISSENSE VARIANTS ON COGNITIVE FUNCTION WITH *DROSOPHILA*

Anna Koutská[1], Tereza Koníková[1], Petra Havlíčková[2], Ivana Kutá-Smatanová[2], Michaela Fencková[1]

[1]Laboratory of Neurogenetics, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic
[2]Department of Chemistry, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic
fenckm00@prf.jcu.cz

Over 200 million people in the world are affected by Intellectual Disability (ID) and/or Autism Spectrum Disorder (ASD), debilitating and often co-occurring neurodevelopmental disorders [1, 2]. They have problems with cognitive and adaptive functioning, including learning, communication, and social skills. ID/ASD have mostly monogenic causes. Majority of the mutations (non-sense, frameshift, or splice-site) are likely gene disrupting (LGD). However, there is a growing number of de novo missense genetic variants with uncertain significance (VUS) that increase the disease risk to similar or even greater degree than LGD [3]. It is challenging to comprehend how VUS affects ID/ASD symptoms; hence it is important to develop an efficient model.

We are introducing 40 conserved recurrent VUS in Drosophila orthologs of ID/ASD genes with CRISPR-HDR. We will investigate their effect on cognitive function with habituation, a conserved form of learning that is based on suppressing a response to a repetitive but meaningless stimulus. Habituation is a prerequisite for higher cognitive functions [4, 5] and, as was shown previously, is affected in Drosophila LGD models of ID/ASD [6]. Thus, habituation is suitable for investigating the effect of VUS on cognitive function in ID/ASD. We use a high-throughput light-off jump habituation platform where the flies suppress their jump response to a light-off stimulus. This study should shed light on the effect of VUS on ID/ASD pathology and, most importantly, on cognitive function that cannot be easily assessed with simpler cell-based models.

1. European intellectual disability research network (2003). 1-129. European Commission.
2. Zablotsky B., Black LI, Maenner MJ etc. (2015) Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014. National Health Interview Survey: National health statistics reports, 87.
3. Iossifov I., O´Roak B. J., Sanders S. J., et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder, Nature, 515(7526), 216-221.
4. Kavšek M. (2004) Predicting later IQ from infant visual habituation and dishabituation: A meta-analysis. JADP, 25(3), 369-393.
5. Domsch H., Lohaus A. and Thomas H. (2009) Prediction of childhood cognitive abilities from a set of early indicators of information processing capabilities. Infant Behavior and Development, 32(1), 91-102.
6. Fenckova M., Blok L.E.R., Asztalos L., et al. (2019) Habituation Learning Is a Widely Affected Mechanism in Drosophila Models of Intellectual Disability and Autism Spectrum Disorders. Biological Psychiatry, 86(4), 294-305.

This work is supported by a grant from the Czech Science Foundation (grant no. 23-07810S) and an EMBO Installation grant (grant no. IG-5310-2023) to M. Fenckova.

SL3

Investigation of crystal structure of galectin-9 in complex with ligands

ZKOUMÁNÍ KRYSTALOVÉ STRUKTURY GALEKTINU-9 V KOMPLEXU S LIGANDY

Michaela Burdová[1], Barbora Kaščáková[1], Michaela Hovorková[2,3], Pavla Bojarová[3], Ivana Kutá Smatanová[1]

[1]Katedra chemie, Přírodovědecká fakulta, Jihočeská univerzita v Českých Budějovicích, Branišovská 1760, CZ-37005 České Budějovice, Česká republika
[2]Katedra genetiky a mikrobiologie, Přírodovědecká fakulta, Univerzita Karlova v Praze, Viničná 5, CZ 12843, Praha 2, Česká republika
[3]Laboratoř biotransformace, Mikrobiologický ústav Akademie věd ČR, Vídeňská 1083, CZ-14200 Praha 4, Česká republika
ivanaks@seznam.cz, burdova.michael@seznam.cz

Galektiny jsou proteiny patřící do skupiny lektinů, které obsahují jednu nebo dvě konzervované domény rozpoznávací sacharidy (carbohydrate recognition domain, CRD), jež váží b-galaktosidové sacharidy [1, 2]. Tyto proteiny byly nalezeny u mnoha mnohobuněčných organismech od bezobratlých až po člověka. Na rozdíl od většiny lektinů nejsou galektiny vázané k membráně, ale vyskytují se jako rozpustné proteiny s intra-i extracelulárními funkcemi, zejména v cytosolu, jádře, extracelulární matrix a v oběhové soustavě. V těle hrají roly v modulaci buněčný procesů. Jako je buněčná morfogeneze, intracelulární transport, signalizace a apoptóza [1, 3].

Galektin-9 (Gal-9) je protein tandemově opakujícího se typu se dvěma CRD. Jeho role v lidském těle je především imunomodulační. Nejvyšší exprese Gal-9 byla zjištěná v tkáních spojených s imunitním systémem (slezina a lymfatické uzliny) a v tkáních endodermálního původu (játra, střeva, žaludek a plíce). Při onemocnění rakovinou dochází v nádorové tkáni ke změně hladiny exprese Gal-9 ve srovnán s nepostiženou tání. Díky této skutečnosti by Gal-9 mohl sloužit jako prognostický marker malignit [2].

Pro určení struktury kompletního proteinu Gal-9, jeho C-domény a N-domény byla použita krystalizační metoda na difúzi par pomocí sedící kapky. Stejnou metodou byla provedena ko-krystalizace za účelem určení přesné polohy vazby ligandu. Pro komplex protein-ligand byly použity komerčně dostupné ligandy thiodigalaktosidu (TDG - VA142 a VA 143). Ko-krystalizace byla provedena jak pro celý protein, tak pro jeho obě domény. V obou případech byly použity komerčně dostupné krystalizační screeny (The Ligand-Friendly Screen, SG1 Screen, PEGRx…).

Získané krystaly byly testovány pomocí JANSi UVEX a následně odeslány na difrakční analýzu na synchrotron BESSY v Berlíně. Vypěstované krystaly C-domény Gal-9 poskytly difrakci s rozlišením 1.6 ?. Dále byly získaný krystaly proteinu Gal-9 vhodné k difrakční analýze. Ko-krystalizací se podařilo vypěstovat i krystaly komplexu C-domény Gal-9 s výše zmíněnými thiodigalaktosidmi. V současné době probíhá další analýza a testování těchto krystalů s cílem potvrdit detailní interakce mezi Gal-9 a sacharidy.

1. Bänfer S., Jacob R. (2022), Galectins. Current biology magazine, 32, pp. 406-408. doi: https://doi.org/10.1016/j.cub.2022.03.065.
2. Heusschen, R., Griffioen, A. W., & Thijssen, V. L. (2013). Galectin-9 in tumor biology: A jack of multiple trades. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1836(1), pp. 177-185. doi:10.1016/j.bbcan.2013.04.006.
3. Arthur, C. M., Baruffi, M. D., Cummings, R. D., & Stowell, S. R. (2014). Evolving Mechanistic Insights into Galectin Functions. Galectins, pp. 1-35. doi:10.1007/978-1-4939-1396-1_1 .

We acknowledge Czech Science Foundation, project no. 22-00262S, and the Ministry of Education, Youth and Sports of the Czech Republic, Mobility Project No. LUC23149.

SL4

A PROGRAM FOR AUTOMATIC CHECKING OF CRYSTAL STRUCTURE SOLUTION RESULTS BASED ON COMPARISON WITH DFT CALCULATION RESULTS

F. Fňukal

University of Chemistry and Technology, Prague, Technická 5, 166 28 Praha 6 – Dejvice
fnukalf@vscht.cz

Introduction

Crystal structure verification based on the comparison with DFT calculation results was already introduced circa 20 years ago [1, 2]. However, only the advancement in computing technology as well as the development in the area of DFT functionals made it possible to perform such calculations on complex organic molecular crystals. Our aim is to develop a program capable of mediating DFT calculations and analysing the results. There already exist commercial pieces of software offering such capabilities, they are however typically fairly expensive. Our aim is therefore also to present a freely available variant of such software.

Crystal structure verification using dispersion-corrected DFT

A DFT calculation uses an experimental structure as an input. During the calculation, the atomic positions and optionally also the cell parameters are optimized in a way that the energy minimum is achieved. The output of a DFT calculation is another structure with a geometry more or less different from the geometry of the experimental structure. The input and output structures can then be compared based on certain selected criteria. These criteria should indicate serious discrepancies in the two structure geometries.

Our implementation – the program checkCIF-DFT

To facilitate performing DFT calculations on crystal structures we developed a program to which we gave the name checkCIF-DFT. An inspiration to us was the web application checkCIF/PLATON [3], which offers consistency and validity checking for experimental crystal structures based on crystallographic diffraction criteria. Our program intrinsically utilizes 3 different DFT programs: Quantum ESPRESSO [4], CASTEP [5] and Orca [6]. Besides that, the molecular mechanics program GULP [7] is also utilized. Our program provides a graphical interface and serves as a mediator between the user and computational programs. Our program can read and visualize data from a CIF file, prepare input files for computational programs, monitor the progress of a calculation and finally, after a calculation has finished, it can analyse the calculation results and point out serious issues.

Input and output structures comparison

To compare the experimental crystal structure and the DFT output crystal structure, it is absolutely essential to choose comparison descriptors that are sufficiently indicative and can therefore reflect serious discrepancies in the compared structures. In our work, we originally used solely the

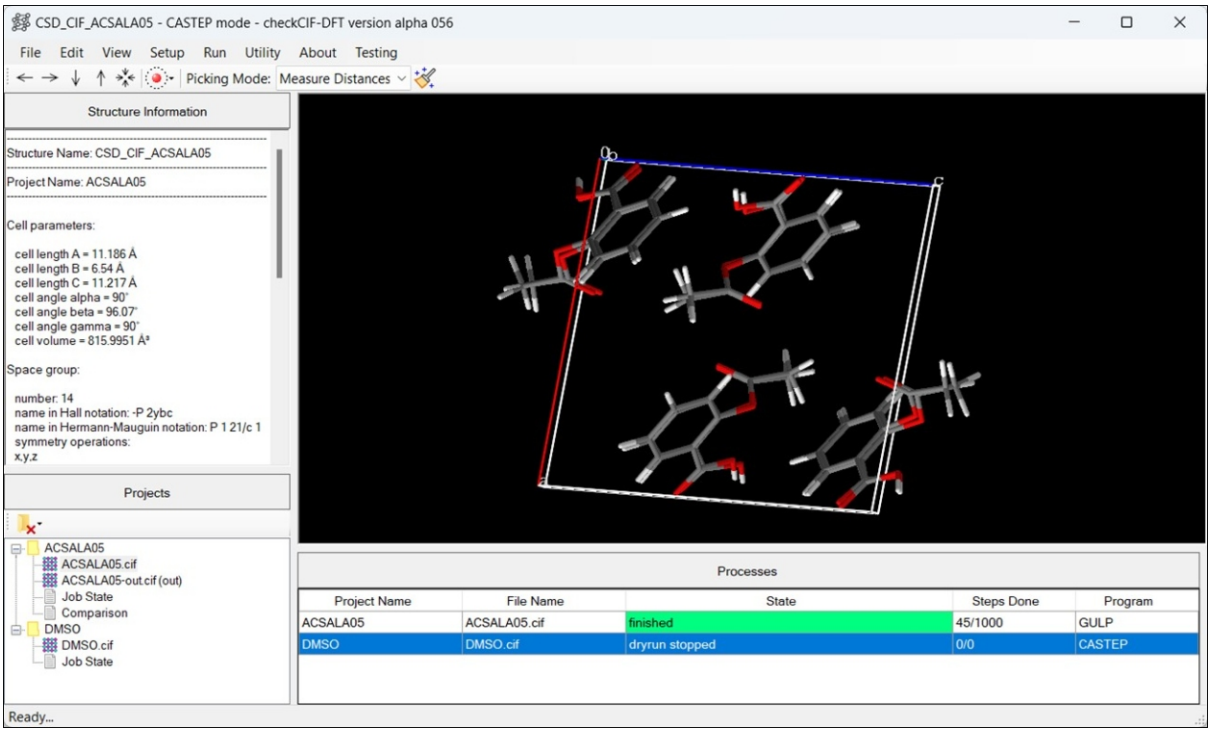

Figure 1. Main window of the program checkCIF-DFT.

descriptor RMSCD developed by other authors [2]. However, as the authors of RMSCD stated themselves, this descriptor doesn't reflect serious issues well enough. For that reason, we decided to include other descriptors. Among the newly implemented descriptors are relative difference in cell volumes, maximal difference in bond lengths, maximal difference in bond angles and others. In our testing so far, we discovered that the tested problematic structures reliably exhibit a serious disagreement in at least one of the used descriptors.

Practical uses of DFT calculation results

DFT calculations can be used for routine verification of experimental crystal structure solutions. Some experimental results may be affected by serious errors due to bad quality of the crystalline sample or other factors. For that reason, a DFT calculation can be useful to assess the trustworthiness of the experimentally obtained data.

Crystal structure prediction represents another field of use for DFT calculations. In such computational experiment a large set of possible crystal structure geometries is generated using lower-level methods (e.g. molecular mechanics). These structures are then refined using the DFT method. The refined set of structures can then be sorted based on the lattice energy, which should reflect the stability of each structure in the set.

Apart from the two examples mentioned above, DFT calculations also find great use in powder diffraction crystal structure solutions. While solving powder diffraction data, a DFT calculation can be used as an intermediate step to achieve a better level of refinement.

DFT method testing

In our work, we've conducted a series of testing calculations to assess how well the DFT method would fare in indicating seriously erroneous crystal structure solutions. In our testing, we chose a set of 5 structures that are known to be fraudulent [8] and a set of 5 structures solved by neutron diffraction experiments, which we deemed to be the most precise and reliable method of determining the crystal structure. For this test we used the CASTEP computational module utilizing the rSCAN functional and MBD dispersion correction. When analysing the results, we concluded that the DFT method together with our improved descriptor system was able to detect that the fraudulent structures were erroneous (Fig. 2).

Conclusions

In our work, we discovered that we were able to detect fraudulent crystal structures using DFT calculations together with an improved system of structure comparison descriptors. The most useful comparison descriptors have shown to be the maximal bond length difference and maximal bond angle difference.

We developed and tested a freely available program that is capable of mediating DFT calculations and analysing the results. This program may help crystallographers in assessing the trustworthiness of crystal structure solutions.

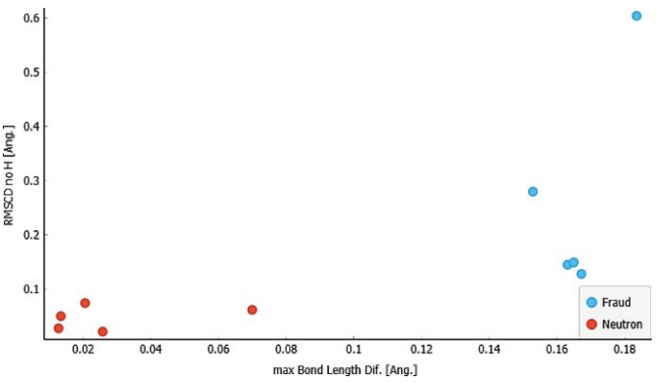

Figure 2. Scatterplot of RMSCD excluding hydrogen atoms against maximal bond length difference for a set of structures solved using neutron diffraction data (red) and a set of structures that are known to be fraudulent (blue).

Table 1. Best 10 trial structures of the compound XXXI from the 7th CSP Blind Test as calculated by the DFT method.

Rank	Structure code	Relative energy [kJ/mol]	Experimental rank
1.	XXXI_structure_59	0.0000	-
2.	XXXI_structure_98	0.7963	1.
3.	XXXI_structure_1	2.0061	2.
4.	XXXI_structure_17	2.5395	-
5.	XXXI_structure_57	2.9341	-
6.	XXXI_structure_34	3.0854	-
7.	XXXI_structure_11	3.1684	-
8.	XXXI_structure_25	3.2349	3.
9.	XXXI_structure_70	3.4779	-
10.	XXXI_structure_20	4.1705	-

1. J. Streek, M. A. Neumann, *Acta .Cryst.*, **B66**, (2010), 544.
2. J. Streek, M. A. Neumann, *Acta .Cryst.*, **B70**, (2014), 1020.
3. (IUCr) IUCr Journals - checkCIF FAQ. https://journals.iucr.org/services/cif/checking/checkfaq.html#what (accessed May 10, 2024).
4. Paolo Giannozzi *et al, J. Phys.: Condens. Matter*, **21**, (2009), 395502.
5. Materials Studio 2023 - CASTEP. https://www.tcm.phy.cam.ac.uk/castep/documentation/WebHelp/CASTEP.html (accessed May 10, 2024).
6. F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2**, (2012), 73.
7. J. Gale, *J. Chem. Soc., Faraday Trans.*, **93**, (1997), 629.
8. *The Lancet*, **375**, (2010), 94.
9. The 7th CSP Blind Test | CCDC. https://www.ccdc.cam.ac.uk/community/ccdc-for-the-community/partnerships-and-initiatives/csp-blind-test/7th-csp-blind-test/ (accessed May 10, 2024).

SL5

APPLICATION OF DFT CALCULATIONS FOR CRYSTAL STRUCTURE VERIFICATION OF PHASES FROM SALT-COCRYSTAL CONTINUUM AREA

S. Chalupná (*maiden name* Šajbanová), M. Hušák

Department of Solid State Chemistry, University of Chemistry and Technology, Prague, Technická 5, Praha 6, 166 28, Czech Republic
sajbanos@vscht.cz

Introductions

Pharmaceutical solid forms such as salts and cocrystals play a crucial role in pharmaceutical applications. The difference between salt and cocrystal is given only by the position of single hydrogen [1], making it essential to develop precise techniques for identifying this position. Differentiation between salt and cocrystal compounds holds significant importance within the pharmaceutical industry, both for regulatory purposes and overall quality control. We are developing a computational method for salt cocrystal differentiation based on DFT (density functional theory) energy calculation. We had already partially tested this method [2]. In this work we confirmed that we can correctly differentiate salt and cocrystal when the H-bond is not extremely strong (reliability rule set by us for H-bond distances longer than 2.613 Å, O–H···N case). Based on the conclusions from the publication we had made a few improvements in our present work: the number of tested structures increased from 95 to 404 and for the main screening the rSCAN functional was used instead of PBE one. Our DFT method optimizes an artificially constructed wrong structure (hydrogen atom placed in salt position near the potential acceptor for cocrystals and vice versa cocrystal position with hydrogen atom placed near the potential donor of the salts). The verification of the method was done based on comparison of the results with an experimentally confirmed correct hydrogen position. 16 cocrystals from the studied set were identified as salt in disagreement with experimental data. These problematic structures were investigated more deeply. We reproduced crystallization and data collection using single-crystal X-ray diffraction (SCXRD) for 7 of them. Complete experimental data were available for 2 problematic structures from the original authors and data re-interpretation was possible. To get the best possible hydrogen positions, we had used for refinement the HAR method as implemented in Olex2 software and NoSpherA2 [3,4,5,6]. We also evaluated whether in these problematic cases more advanced functionals (r2SCAN, PBE0, PBE50) could provide results consistent with experimental data.

Table 1. Results of calculation on 404 structures from the zone with -1 pKa 4 (RSCAN fine + MBD).

Pure cocrystal	Pure salt	Salt-cocrystal continuum phase
301	16	87

Methods

The DFT calculations were performed using CASTEP code [7]. Since the cell parameters were assumed to be accurately obtained experimentally, we solely performed only optimization of atomic positions. We had used rSCAN functional with MBD dispersion correction and automatic fine basis precision [8,9]. The data were prepared in checkCIF-DFT software [10]. The optimization was always performed from both artificial salt and cocrystal starting models. Computation was performed on Karolina supercomputer at TU Ostrava, Czech Republic.

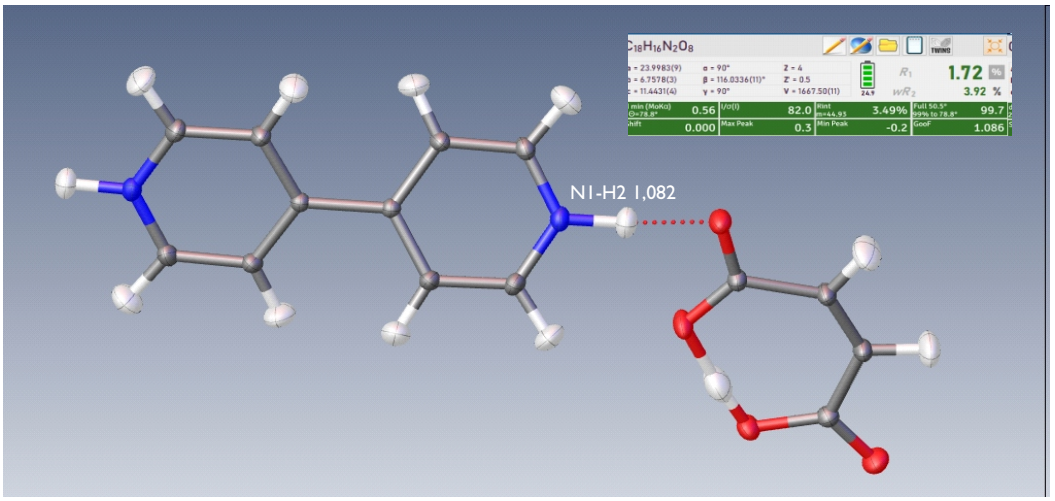

Figure 1. Structure of 4,4'bipyridine and maleic acid (GIPQAX) in Olex2 refined by HAR method with hydrogen atoms treated as anisotropic. The structure was originally experimentally solved incorrectly as cocrystal.

For cases where we have crystallized the structure or re-interpreted the data of the original authors, we used HAR method as implemented in Olex2 software and NoSpherA2 module [3,4,5,6]. For the wavefunction calculation we had used def2-TZVP localized base, r2SCAN functional and Orca 5.0 software [11]. The refinement was in all cases done by two methods. The first method was based on refinement of the problematic hydrogen in single position. The second method was based on refinement of this hydrogen in two positions as disordered one. The donor and acceptor distances to the hydrogen were in the second case restrained to the value 0.95 Å with weight corresponding to 0.01 Å esd. For final CIF deposition, the refinement results based on the first method were used only because we believe the disorder model does not correctly reflect the real state of the phases.

Results

We had confirmed a correct cocrystal structure determination in 301 cases. For 87 structures we had identified that the phase determination is suspicious, and the structures probably create a salt-cocrystal continuous phase. This phases behaviour will be described in separated study.

From the 16 phases exhibiting consistent salt behaviour by our methodology, we experimentally proved that 2 are true salts (OGEPIA, GIPQAX). We believe that 3 others (ODOHIZ, CITSAZ10, VODCOH) were incorrectly solved by the original authors, and our DFT method using rSCAN functional correctly identified these as salts. For the 4 structures that violate the reliability rule we established in the previous article (DFT method can correctly differentiate salt and cocrystal when the H-bond is not extremely strong; H-bond distances longer than 2.613 Å, O–H···N case), we confirmed the DFT methodology based on rSCAN functional works correctly and the problem is with the experimental structure determination. In 9 cases we confirmed that the DFT method based only on rSCAN functional is not reliable and unsuitable for cocrystal/salt distinguishing with strong H-bond. However, advanced functionals (r2SCAN, PBE0, PBE50) can correct these discrepancies in some cases. For future prediction we suggest for the salt-cocrystal differentiation the r2SCAN functional which provides correct results for O–H···N bonds longer than 2.554 Å, compared to our previous 2.613 Å limit. The computational cost of r2SCAN is comparable to rSCAN, making it suitable for large-scale screening.

1. Aitipamula, S.; Banerjee, R. Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. Q. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Thaper, R. K.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J., Polymorphs, Salts and Cocrystals: What's in a Name? (vol 12, pg 2147, 2012). Cryst. Growth Des. 2012, 12 (8), 4290-4291.

2. Hušák, M.; Šajbanová, S.; Klimeš, J.; Jegorov, A. (2022) Acta Crystallographica Section B. 78, 5, 781 – 788.

3. Jayatilaka & Dittrich., Acta Cryst. 2008, A64, 383.

4. Capelli et al., IUCrJ. 2014, 1, 361.

5. Kleemiss et al., Chem. Sci. 2020, DOI: 10.1039/D0SC05526C.

6. Midgley et al., Acta Cryst. 2021, A77, 519-533.

7. Clark, S. J. , Segall, M. D. , Pickard, C. J. , Hasnip, P. J. , Probert, M. J. , Refson, K & Payne M. C. (2005) Zeitschrift fuer Kristallographie 220(5-6), 567-570.

8. Bartók, A. P. & Yates, J. R. (2019). J. Chem. Phys. 150, 161101-1-161101-5.

9. Hermann, J., & Tkatchenko, A. (2018) J. Chem. Theory Comput. 14(3), 1361–1369. doi:10.1021/acs.jctc.7b01172.

10. Fňukal, F Automatická kontrola výsledků řešení krystalové struktury na základě srovnání s výsledky DFT výpočtu., Vysoká škola chemicko-technologická v Praze, 6.6. 2024.

11. Neese, F. (2022). Software update: The ORCA program system—Version 5.0. Wiley Interdisciplinary Reviews. Computational Molecular Science, 12(5). https://doi.org/10.1002/wcms.1606.

This work was supported by the grant of Specific university research—grant No. A2_FCHT_2024_054.

L8

THE POTENTIAL OF DISPERSION-CORRECTED DENSITY FUNCTIONAL THEORY CALCULATIONS TO VERIFY CSD DATABASE CONTENT

M. Hušák

University of Chemistry and Technology, Prague, Technická 5, 166 28 Praha 6 – Dejvice
husakm@vscht.cz

Introduction

The idea to verify crystal structure by comparison with DFT calculation results was introduced already 20 years ago [1,2]. Due to advances in the computation technologies, DFT functionals development as well as bit problematic results of the original work we had chosen to refresh the whole idea. The target of our work is to develop this methodology and to test whatever verification of the complete CSD content by this method is eventually possible.

Improvement of the methods

The original verification method [1,2] was primary based on the experimental and DFT results comparison by Cartesian displacement (RMSCD) descriptor only. RMSCD is de-facto a RMSD modified to be able to compare atomic positions in different unit cells. It was already mentioned [1] this descriptor was not able to clearly separate totally artificial fraud structures from correctly solved one. We had tested several other descriptors to separate problematic structures from correct ones. Our test indicates the maximal difference in bond distance and maximal difference in bond angle can better indicate problematic structures than RMSCD, see Fig. 1. The artificial fraud structures are clearly separated from correct results.

Another improvement of the method is the use of modern meta-GGA functional (r2SCAN) with up-to-date dispersion correction (MBD) instead of the PBE functional and first generation of the Grimme dispersion correction as used in the original work. This method of energy calculation leads to more realistic atom positions and lattice parameters than the original one. The whole methodology was implemented in a form of checkCIF-DFT software which is a subject of another presentation.

Test on 100 semi-randomly selected structures from CSD

In the first tests we had extracted from CSD 194017 structures by following criteria: published after 2013, non-disordered pure organic structures, no errors detected by CSD, solved from single crystal. We had used only pure organic structures because metal presence can generate problems related to open shell and not clear spin state which is hard to handle automatic for DFT calculation. From the mentioned structures we had selected in a semi-random way 100 one (with one addition criteria - original diffraction data must be deposited). For 30 structures it was impossible to perform the verification - see Tab. 1 for reasons.

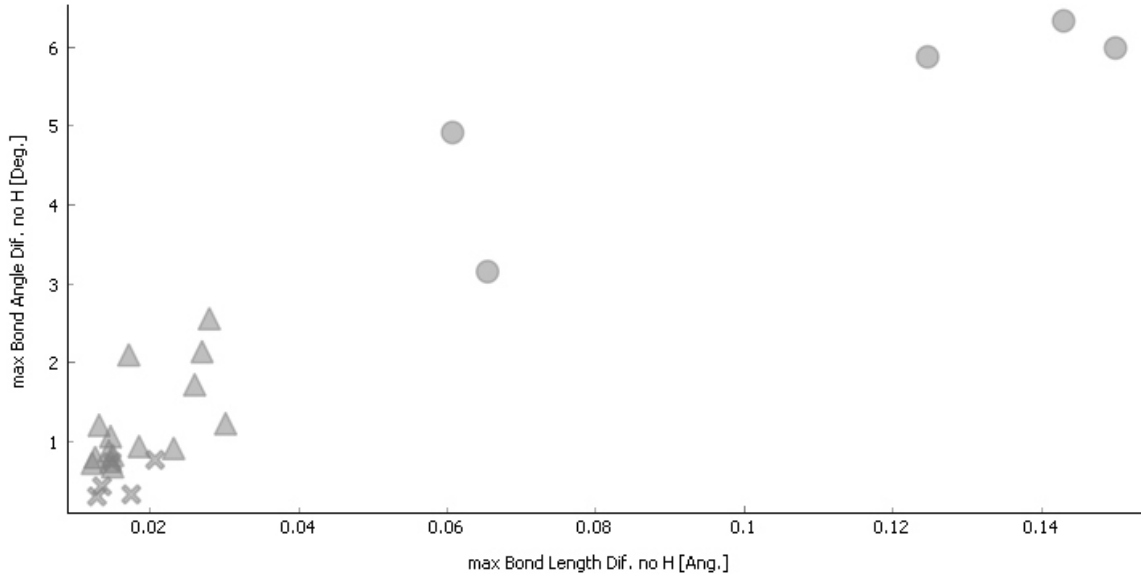

Figure 1. Maximal bond length difference versus maximal bond angle difference (hydrogen atoms excluded). × - correct neutron structures, - correct X-ray structures, ○ - artificial fraud structures

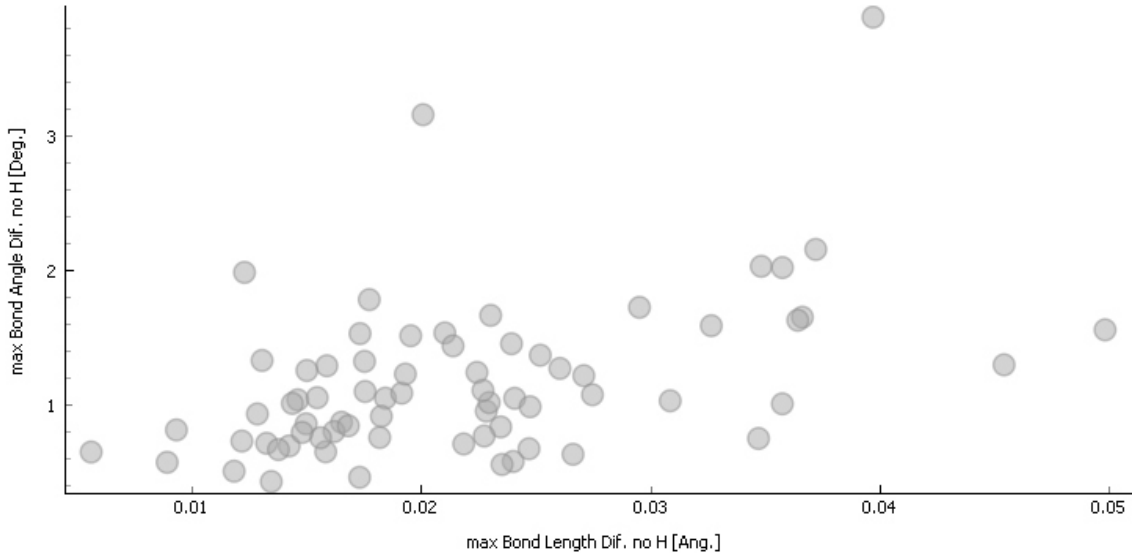

Figure 2. Maximal bond length difference versus maximal bond angle difference (hydrogen atoms excluded) for the 70 fully processed structures.

The verification process was fully finished for 70 structures from the test set. The maximal bond and angle difference descriptor visualisation graph is on the Fig 2.

Pre filtering test on the bigger structures sample

Based on the issues with the 100 structures sample, we had tried to run some simpler test on the whole 194017 set. Eventual computational non-expensive pre-filtering can help to save computational resources as well. The inspiration of the test was the PLATON/checkCIF code.

The first test was a test for correct space group determination. The test checks whatever there exist a higher symmetry able to describe the structure or whatever the structure is not described by super cell. The test was done by the help of Spglib[3] library in a similar way as it is done by ADDSYM code in PLATON. An issue was detected for 622 structures (0.32%).

Another test was done for Solvent Volume presence calculation in the unit cell. The Solvent Volume was calculated as suggested in the BYPASS article [4] so it corresponds exactly to values calculated by PLATON/checkCIF. 20 057 structures (10.34 %) with Solvent Area > 40 Å^3 was detected.

The last test was done by checking the correspondence between `chemical_formula_sum` information in the CIF file and formula generated from tom coordinates. A disagreement was found for 45 974 structures (23.70%). In most often cases the disagreement is the result of generating more than asymmetric unit cell atoms by CSD ConQuest to get a complete molecule. Unfortunately, this data cannot be used directly for DFT calculation without pre-processing. In a lot of cases the results indicate missing atoms in the structure as well.

Table 1. Issues detected during the CSD 100 structures test

Issue description	Number of structures
Incorrect space group	1
Missing disordered on nonsense hydrogen atoms positions.	4
Voids indicating missing solvents (Solvent Area > 40 Å^3)	12
To big structures (performance issues)	2
Not converging after 100 optimization steps	7
Not fitting in used computer memory	1
Duplicated atoms generated over symmetry	3
Correctly calculated structures	70

Conclusions

DFT method can reliably detect issues with incorrectly determined structures. Unfortunately, its use is limited by computational resources, presence of metals in the structure and structure disorder occurrence. Only a sub-set of CSD can be checked by this methodology due to multiple issues with deposited structures. Even on a small 70 structures sample from CSD, structures with symptoms similar to fraud structures (compare Fig 1 and Fig 2) can be detected. A test made on large sample by the use of supercomputer, or a test done by alternative methods (ML force field) is required for more conclusions.

1. J. Streek, M. A. Neumann, *Acta .Cryst.*, **B66**, (2010), 544.
2. J. Streek, M. A. Neumann, *Acta .Cryst.*, **B70**, (2014), 1020.
3. Spglib : https://arxiv.org/abs/1808.01590.
4. P. Sluis, A. L. Speak, *Acta .Cryst.*, **A46**, (1990), 194.