Available online at www.sciencedirect.com

ScienceDirect

Mendeleev Commun., 2022, **32**, 735–738

Mendeleev Communications

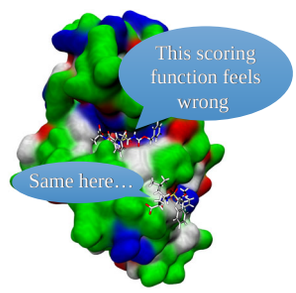
Reducing false-positive rates in virtual screening *via* cancellation of systematic errors in the scoring function

Michael G. Medvedev,*[a] Oleg V. Stroganov,[b] Artem O. Dmitrienko,[a,c] Maria V. Panova,[a] Alexey A. Lisov,[a] Igor V. Svitanko,[a,d] Fedor N. Novikov*[a,d] and Ghermes G. Chilov[e]

[a] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation. E-mail: medvedev.m.g@gmail.com*
[b] *BioMolTech Corp., Toronto, Ontario M2L 1L1, Canada*
[c] *Department of Chemistry, M. V. Lomonosov Moscow State University, 119991 Moscow, Russian Federation*
[d] *National Research University Higher School of Economics (HSE University), 101000 Moscow, Russian Federation. E-mail: fnovikov@hse.ru*
[e] *Molecular Technologies, LLC, Skolkovo Innovation Center, 121205 Moscow, Russian Federation*

DOI: 10.1016/j.mencom.2022.11.009

Here we propose an over-the-hood docking method that compensates for systematic errors in the docking force fields. This method explicitly estimates the interaction energy of the ligand with the protein surface and uses it as a baseline to estimate the actual binding energy in the active site. It improves the accuracy of virtual screening in the LeadFinder package by up to 48%.



Keywords: docking, virtual screening, protein–ligand interactions, thymidylate synthase, RNase, ribonuclease, PARP.

Evaluation of protein–ligand interactions is a key technology in drug discovery, and virtual ligand screening is one of the practical methods[1] commonly used in conjunction with various experimental methods. Despite concerns about the cost and accuracy of currently available wet-lab techniques, virtual screening (VS) is still far from replacing experimental methods.[2] The fundamental reason for this is that current approaches do not always reliably predict the free energy of protein–ligand binding, so the output of modeling techniques fails to satisfy researchers.[3]

The search for novel ideas to improve the accuracy of structure-based VS methods continues. In particular, performances of VS can be improved by structural or pharmacophore filtration,[4,5] as well as SIFt (Structural Interaction Fingerprints)[6] and IFP (Interaction Fingerprints).[7] These and related approaches exclude ligands exhibiting unusual binding patterns, thereby limiting comparison to ligands having similar poses in the active site of the target protein. This leads to favorable error cancellation which masks defects in the scoring function. For example, if the scoring function systematically overestimates van der Waals interactions, the magnitude of the overestimation will be similar among related ligands with similar binding patterns and will not affect the ranking as much as without structural filtration.[8,9] The LeadFinder system shows good results in combination with the FEP (Free Energy Perturbation) setting, which allows taking into account conformational filtration in the ligand–target coordination act, as well as the effect of the solvent on the parameters of the resulting complex.[10–12]

Structural filtration approaches require a significant amount of knowledge on target protein binding with various active ligands. For this reason, they are unreliable for novel targets. Another obvious shortcoming of these approaches rests in their inability to detect active ligands with unprecedented binding patterns. Therefore, to improve existing protocols, a method is needed that is simple enough for high-throughput screening of molecular libraries and yet capable of utilizing error cancellation to accurately rank ligands with novel binding patterns.

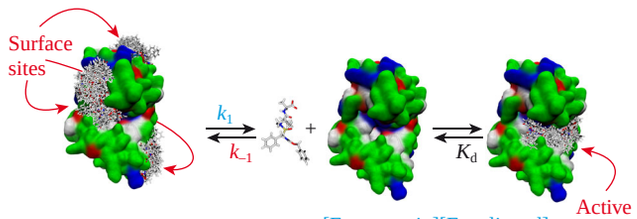
Here, we present a new 'On-Top Docking' (OTD) approach, an add-on to the LeadFinder system, which explicitly estimates the energy of ligand interaction with the protein surface and uses it as a baseline for assessing the actual binding energy to the active site. Since biases in the scoring function would manifest themselves in the binding energy to both the active site and the protein surface, one would expect the difference between them to be a much better predictor of ligand activity than the uncorrected binding energy in conventional docking.

Previously, the Hammerhead[13] and BINDSURF[14] programs have utilized automatic assessment of ligand binding to various surface sites of the protein, but both of them were aimed at finding the actual active site of the protein, rather than assessing non-specific ligand binding. We have found that the OTD approach, when used with the LeadFinder system, significantly improves the VS results of 35 test proteins (in particular, those with open active sites and non-standard net charges) and is compatible with various scoring functions.

Error cancellation in building an OTD model can be achieved using the energy of a non-specific ligand binding to a random protein site as a baseline. Using the protein's own surface as a 'random protein site', we arrive at the following model for the corrected docking score:

$$\mathrm{SCORE_{OTD}} = \mathrm{SCORE_{AS}} - \mathrm{SCORE_{SS}}. \quad (1)$$

© 2022 Mendeleev Communications. Published by ELSEVIER B.V. on behalf of the N. D. Zelinsky Institute of Organic Chemistry of the Russian Academy of Sciences.



Conventional model: $k_1 >> k_{-1}$; $K_d = \frac{\text{[Free protein][Free ligand]}}{\text{[Active site-bound ligand]}}$

OTD model: $k_1 << k_{-1}$; $K_d = \frac{\text{[Surface-bound ligand]}}{\text{[Active site-bound ligand]}}$

Figure 1 Possible physical interpretation of equation (1). The equations for the dissociation constants are given for the conventional (used in conventional docking) and OTD models.

Here $SCORE_{AS}$ and $SCORE_{SS}$ are the scores computed from all docking runs into the active site and to the surface sites, respectively. Since in conventional docking, ligands dock 'inside' the protein into its active site, we denote the procedure, which takes into account the affinity of the ligand to the protein surface, as 'On-Top Docking', or OTD for short.

The use of the protein's own surface additionally allows one to exclude ligands that are not active due to the high affinity for the protein surface and leads to a simple physical interpretation of equation (1), which corresponds to the difference between the energies of the ligand bound at the active site and on the protein surface (Figure 1).

Optimal parameters for OTD were determined as following:

To make our algorithm as self-contained as possible, the surface sites were centered on 25 randomly selected atoms uniformly distributed over the protein surface [Figure 2(*a*), golden spheres].

By performing 10 runs into each of the 25 surface sites, we get 250 docking scores, which should be tallied into $SCORE_{SS}$ from equation (1). Since docking runs frequently fail to find a proper ligand pose, thus resulting in huge outliers, averaging the scores has little sense, and ranking statistics is required to select an appropriate value from 250. We have decided to split this problem into two parts: first, we chose the run into each surface site, and then we choose the appropriate value from the remaining 25.

We have considered four possible choices among the runs into each surface site: the minimum, median and maximum energies, as well as an arbitrary one (Figure 2). The latter simulates carrying a single docking run into each surface site.

For surface sites, we have considered centiles (see Figure 2). Due to the automatic procedure, some of the surface sites could accidentally get into the active site, while some of the others could be under the surface of the protein. Thus, the value of interest should not be within the lowest or highest scores, and likely is close to the median.

In accordance with equation (1), the $SCORE_{SS}$ is then subtracted from the $SCORE_{AS}$ (minimum over 10 runs into active site), and the resulting value is used as a ligand score to compute AUROC for the given target. The dependence of the AUROC, averaged over all targets, on the selected on-top runs and on-top site score quantiles is shown in Figure 2(*b*).

Notably, the maximum score curve [Figure 2(*b*), solid line] has a wide horizontal plateau at the top ranging from ~0.25 to ~0.75 quantiles, indicating that 25 surface sites are superfluous. On contrary, the minimum score curve [Figure 2(*b*), dot-dashed line] has an inclined plateau, showing that the minimum scores somewhat overestimate the required correction; although they still lead to a moderate improvement over the conventional docking [Figure 2(*b*), black horizontal line]. The curves for the median scores [Figure 2(*b*), dashed line] and arbitrary scores [Figure 2(*b*), dotted line] almost coincide and provide a considerable improvement over

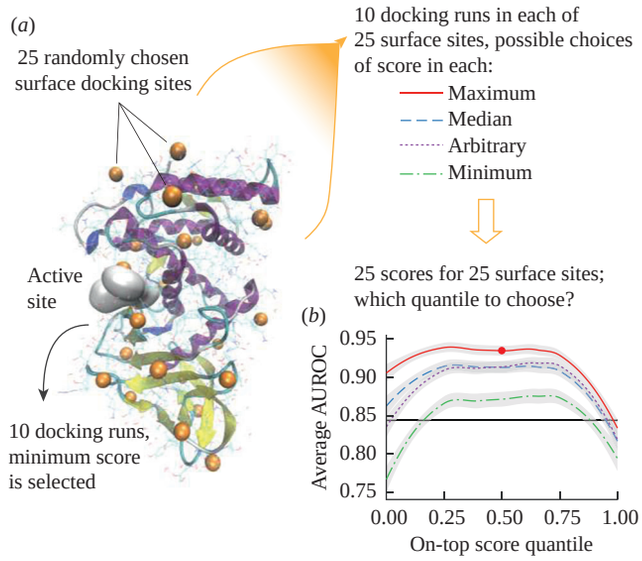


Figure 2 OTD concept and selection procedure. (*a*) An example of a protein with an active site (silver surface) and 25 randomly selected surface (on-top) docking sites (golden spheres). (*b*) Dependence of the average AUROC value on the score, selected for each surface site from minimum (dot-dashed), arbitrary (dotted), median (dashed) or maximum (solid) energies, and quantiles of scores across all on-top sites. The gray areas represent the 75% confidence intervals of the curves. The OTD procedure is denoted with the red dot. The black horizontal line indicates the performance of the conventional docking.

the results for the minimum scores. The BEDROC and Enrichment factor metrics provide identical patterns to AUROC (see Figure S1).

Thus, the best VS accuracy for our dataset (all 35 proteins) was obtained using the procedure given by equation (2). Here, S_{AS} and S_{SS} are the scores obtained in individual runs of docking into the active site and surface sites, respectively.

$$SCORE_{OTD} = \min_{\text{runs}}(S_{AS}) - \text{median}_{\text{surface sites runs}}[\max(S_{SS})] \quad (2)$$

Below we refer to this procedure as OTD. Thus, OTD requires 10 docking runs into the protein active site (as in conventional docking) and an additional 10 runs into each of 25 random sites on the protein surface; the resulting score is calculated as the difference

Table 1 Comparison of the performances of the OTD procedure for proteins and the conventional procedure with low accuracy.[a]

PDB ID	Protein type	Conventional docking accuracy (%)	OTD accuracy (%)	Improvement[b] (%)
1qhc	Ribonuclease	44.6	92.9	48.3
1efy	Transferase	46.7	68.0	21.3
1f4g	Transferase	58.0	93.9	35.9
1rnt	Ribonuclease	62.7	97.7	34.9
1eve	Hydrolase	68.8	81.3	12.6
1uy6	Hydrolase	70.4	81.0	10.6
1eix	Lyase	76.1	95.3	19.1
1e66	Hydrolase	78.1	85.3	7.2
1m2z	Nuclear receptor	79.1	82.8	3.7
2qwg	Hydrolase	81.0	93.6	12.5
1cx2	Oxidoreductase	82.0	87.1	5.2
1kim	Kinase	84.0	94.7	10.7
1o86	Hydrolase	84.1	94.8	10.7
1bxo	Protease	84.5	95.2	10.7
2oh4	Kinase	84.6	91.9	7.3

[a] The full set of proteins is presented in Tables S1 and S2, and ROC curves are shown in Figure S2 (see Online Supplementary Materials). [b] Difference between OTD accuracy and conventional docking accuracy.

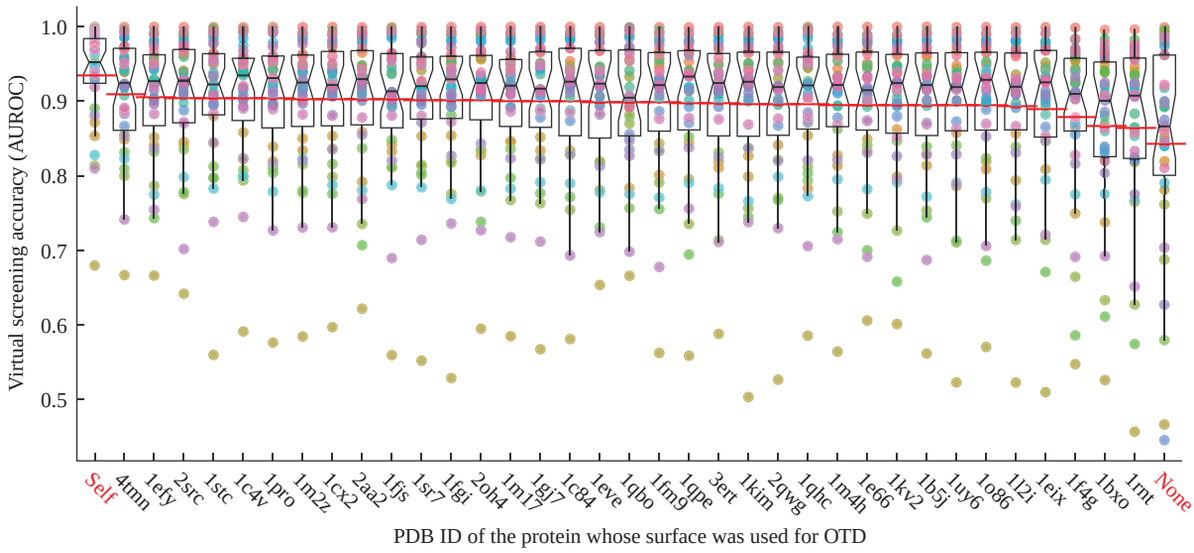


Figure 3 Effect of different protein surfaces on the OTD performance. Dots indicate individual proteins (each protein has its own unique color), red dashes are average AUROCs over all proteins.

between the lowest score in the active site and the median of the highest scores in the surface sites.

Equation (2), obtained for $SCORE_{OTD}$, lacks a clear physical meaning due to the use of the maximum (*i.e.*, worst) score over all runs into a surface site. Therefore, it is unlikely to represent an actual physical process (shown in Figure 1), confirming the role of error cancellation. Nevertheless, it has a pure practical sense: if the worst scores in more than half of the surface sites are comparable to the best score in the active site, the ligand is unlikely to be active, regardless of its affinity for the active site.

The results obtained indicate the efficiency of the proposed procedure for predicting the activity of ligands. Still, there are two possible explanations: (1) canceling the systematic errors of the scoring functions and (2) accounting for actual ligand–surface binding (see Figure 1). In the first case, the target's own surface should not differ from the surface of any other protein. To answer this question, we cross-combined each active site of the protein with the surfaces of each protein under consideration and compared

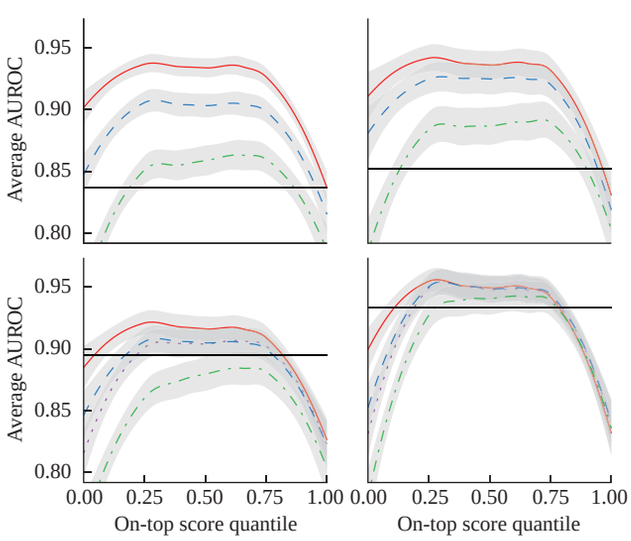


Figure 4 Dependence of the average AUROC value on the score selected for each surface site from minimum (dot-dashed), arbitrary (dotted), median (dashed) or maximum (solid) energies and quantile of scores across all on-top sites for two scoring functions, dG (top) and VS (bottom). The dependencies are shown for two subsets of target proteins, not used to parameterize the VS score (left) and used for this task (right). The gray areas represent the 75% confidence intervals of the curves. Black horizontal lines indicate performances of the conventional docking.

the results with the original OTD using the protein's own surface ('Self') and conventional docking ('None'). The docking settings were identical to those described above, except that we performed only three runs into each surface site.

It turned out, that while the surfaces of all proteins lead to a noticeable improvement compared to the conventional docking, the protein's own surface is specific and gives even more outstanding results (Figure 3). This indicates that either the constructed approach benefits from the physical sense of equation (1) (*i.e.*, the affinity of ligands to the protein surface affects their activity), or that the best error cancellation is achieved when the surface electrostatics is similar to the active site electrostatics (see below for the dependence of docking accuracy on protein net charge).

The results presented above show that the proposed approach leads to a vast improvement in VS accuracy when used with the dG scoring function. To test its versatility, we used OTD along with the VS scoring function, which is another parameterization of the dG score. While dG score was designed to accurately predict ligand binding energies, the VS score was specifically optimized for virtual screening. Its training set consisted of 16 proteins, each of which is included in our test set. For the latter reason, it is instructive to study the improvements on the two protein subsets separately (Figure 4).

According to Figure 4, OTD has the same optimal parameters corresponding to equation (2) for all four combinations of two scoring functions and two protein subsets. As expected, the VS score accuracy (average AUROC) for the subset of proteins used in its development is extremely high (93%), but OTD further improves this to 95%. The accuracy of the VS score on the second subset is 90% and becomes 92% with the OTD. Thus, OTD is a generally applicable approach compatible with different scoring functions.

Even more important for drug discovery is the question of whether a given protein requires OTD for obtaining accurate results. An analysis of the influence of OTD on virtual screening accuracy on the studied protein showed that correcting systematic errors in the scoring function is vital for proteins with open active sites[15] and unusual net charges[16] (Figure 5).

Open active sites obviously have a weaker ability to select suitable ligands based on their geometry, making them more prone to scoring function errors, exacerbating electrostatic or hydrophobic interactions. The relatively good performance of conventional docking for proteins with a net charge of −10 to +10 indicates that the scoring function has a more balanced performance

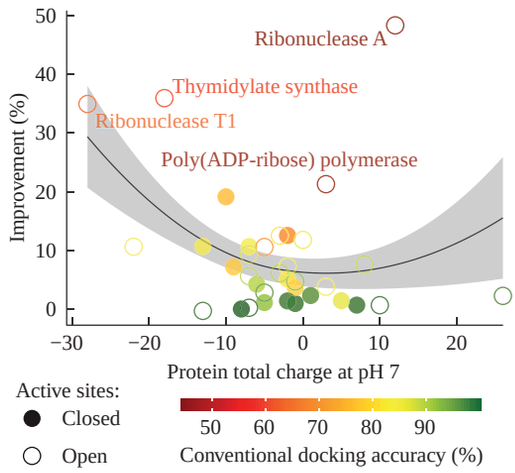


Figure 5 Improvement in docking accuracy when moving from conventional docking to OTD as a function of protein net charge at pH 7 (estimated with BuildModel[17]) and active site type. Each empty or filled circle corresponds to a protein with an open or closed active site, respectively, and is colored according to the accuracy of conventional docking with this protein. The black curve represents the moving average with the gray area denoting the 75% confidence interval.

for proteins with abundant charges. Fortunately, the OTD approach balances the performance of the scoring function across all protein charges.

The proteins most sensitive to biases in scoring functions are thymidylate synthase, RNase A, RNase T1 and poly(ADP-ribose) polymerase, all of which are promising targets for drug development and are known to be hard tasks for the conventional docking (for details, see Online Supplementary Materials). OTD improves the accuracy of virtual screening for these targets by 36%, 48%, 35% and 21%, respectively.

In conclusion, we have demonstrated that systematic errors in scoring functions for protein–ligand binding energy can be canceled out by subtracting the energy of the same ligand bound to a random protein surface computed using the same scoring function. The best results are achieved when the protein's own surface is used and the binding energy is calculated as the median of the maximum energies computed in several surface sites. We refer to this method as OTD. For a comparison with other methods, see the cited paper.[18]

This approach significantly reduces the high false-positive rates of structure-based VS observed in many studies: compounds whose binding energy is exaggerated by the scoring function appear to be active towards almost any protein, but are also predicted to interact strongly with a random protein surface, so the OTD corrects their ranking.

The devised approach significantly improves docking accuracy for all 35 tested proteins. The greatest improvements are observed for proteins with unusual net charges and open active sites. In addition, we found that accounting for systematic errors in the scoring function is essential for some important drug targets: thymidylate synthase, ribonuclease A, ribonuclease T1 and poly(ADP-ribose) polymerase.[19]

Online Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi: 10.1016/j.mencom.2022.11.009.

References

1 J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth and J. J. Irwin, *Nature*, 2019, **566**, 224.
2 A. B. Richon, *Drug Discovery Today*, 2008, **13**, 665.
3 G. L. Warren, C. Webster Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff and M. S. Head, *J. Med. Chem.*, 2006, **49**, 5912.
4 D. Muthas, Y. A. Sabnis, M. Lundborg and A. Karlén, *J. Mol. Graphics Modell.*, 2008, **26**, 1237.
5 F. N. Novikov, V. S. Stroylov, O. V. Stroganov and G. G. Chilov, *J. Mol. Model.*, 2010, **16**, 1223.
6 Z. Deng, C. Chuaqui and J. Singh, *J. Med. Chem.*, 2004, **47**, 337.
7 G. Marcou and D. Rognan, *J. Chem. Inf. Model.*, 2007, **47**, 195.
8 L. V. Romashov, A. A. Zeifman, A. L. Zakharenko, F. N. Novikov, V. S. Stroilov, O. V. Stroganov, G. G. Chilov, S. N. Khodyreva, O. I. Lavrik, I. Yu. Titov and I. V. Svitan'ko, *Mendeleev Commun.*, 2012, **22**, 15.
9 V. S. Stroylov, T. V. Rakitina, F. N. Novikov, O. V. Stroganov, G. G. Chilov and A. V. Lipkin, *Mendeleev Commun.*, 2010, **20**, 263.
10 P. V. Rusina, I. Yu. Titov, M. V. Panova, V. S. Stroylov, Y. R. Abdyusheva, E. Yu. Murlatova, I. V. Svitanko and F. N. Novikov, *Mendeleev Commun.*, 2020, **30**, 430.
11 O. V. Stroganov, F. N. Novikov, M. G. Medvedev, A. O. Dmitrienko, I. Gerasimov, I. V. Svitanko and G. G. Chilov, *J. Comput.-Aided Mol. Des.*, 2020, **34**, 121.
12 V. S. Stroylov and I. V. Svitanko, *Mendeleev Commun.*, 2020, **30**, 419.
13 W. Welch, J. Ruppert and A. N. Jain, *Chem. Biol.*, 1996, **3**, 449.
14 I. Sánchez-Linares, H. Pérez-Sánchez, J. M. Cecilia and J. M. García, *BMC Bioinf.*, 2012, **13** (Suppl 14), S13.
15 A. N. Jain, *J. Comput.-Aided Mol. Des.*, 2008, **22**, 201.
16 E. Gianazza and P. Giorgio Righetti, *J. Chromatogr. A*, 1980, **193**, 1.
17 O. V. Stroganov, F. N. Novikov, A. A. Zeifman, V. S. Stroylov and G. G. Chilov, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 2693.
18 I. Yu. Titov, V. S. Stroylov, P. V. Rusina and I. V. Svitanko, *Russ. Chem. Rev.*, 2021, **90**, 831.
19 I. R. Grin, G. V. Mechetin, R. D. Kasymov, E. A. Diatlova, A. V. Yudkina, S. N. Shchelkunov, I. P. Gileva, A. A. Denisova, G. A. Stepanov, G. G. Chilov and D. O. Zharkov, *Molecules*, 2021, **26**, 6668.

Received: 3rd June 2022; Com. 22/6924