Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Structure of the 5' untranslated region in SARS-CoV-2 genome and its specific recognition by innate immune system via the human oligoadenylate synthase 1†

Emmanuelle Bignon,*[a] Tom Miclot,[a,b] Alessio Terenzi[b], Giampaolo Barone[b], and Antonio Monari*[a,c]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The 2-5-oligoadenylate synthetase 1 (OAS1) have been identified as one of the key enzymes driving the innate immune system response to SARS-CoV-2 infection and has been related to COVID-19 severity. OAS1 is a sensor of endogenous RNA that triggers the 2'-5' oligoadenylate/RNase L pathway in response to viral infections, ultimately activating the RNA-Lyase which cleaves endogenous and exogenous RNA hence impeding the viral maturation. Upon SARS-CoV-2 infection, OAS1 is responsible for the recognition of viral RNA and has been shown to possess a particularly high sensitivity for the 5'-untranslated (5'-UTR) RNA region, which is organized in a double-strand stem loop motif (SL1). Yet the structure of the nucleic acid/protein complex has not been resolved. Here, we report the structure of the OAS1/SL1 complex generated by molecular modeling, including enhanced sampling approaches. We also pinpoint how SL1 region enhances the interaction network with the enzyme, promoting specific hydrogen bonds which are absent in normal double strand RNA fragments, hence rationalizing the high affinity shown by OAS1.

Upon sensing of endogenous double-stranded RNA, OAS1 catalyzes the formation of the secondary messenger 2'-5' oligoadenylate, which subsequently activates the RNase L enzyme responsible for RNA cleavage, thus stopping viral replication. As a matter of fact, OAS1 plays an important role in the response to SARS-CoV-2 infection. Indeed, the severity and outcome of COVID-19 have been linked to OAS1 polymorphisms[1–3], making it an interesting target for antiviral drugs development[4,5]. OAS1 is sensitive to RNA sequence[6], and has been proposed to have a strong affinity for the 54 first nucleotides of the 5' untranslated region (5'-UTR) of SARS-CoV-2[2]. The characterization of the secondary structure of this RNA region indicates that it is organized in two distinct stem loops, of whom the first one (SL1) exhibits a remarkably high affinity for OAS1[7,8]. Furthermore, the SL1 structure appears to have a crucial role in regulating the genome replication, and particularrly the action of the RNA dependent RNA polymerase.[9,10] Models of the SL1 structure have been reported, based on secondary structure predictions[11,12], and more recently on experimental data[13]. However, there is no tertiary structure of the SL1 ds-RNA in interaction with OAS1, which would provide crucial information for antiviral drug design[14]. Here, we describe the tertiary structure of the OAS1/SL1 complex as well as its dynamical behavior resolved using all-atom equilibrium molecular dynamics (MD) simulations and their Gaussian Accelerated extension (GAMD). The characterization of the interaction network between SL1 and OAS1 highlights the importance of the hairpin organization in promoting the high affinity of the immune system protein for this specific RNA fragment. Hence, our work may provide important molecular basis for antiviral drug development, specifically acting against SARS-CoV-2 infection and targeting the OAS1-RNAse L pathway. The possibility of targeting the UTR genome regions is also attractive due to the fact that this region appears fundamental to finely regulate RNA replication.[15]

Based on the reported secondary structure of the 5'-UTR sequence of SARS-CoV-2, the SL1 region that has been proposed to interact with OAS1 was generated using Unafold web server.[16] Our 3D model, spanning the first 40 residues of the 5'-UTR sequence, shows the spatial organization of the rG7-rC33 SL1 stem loop harboring a characteristic central bulge (residues rA12 and rA27-rC28) which connects a double stranded region to a loop (residues rU18-rC21) at its extremity - see Figure 1.
Unrestrained μs-scale MD simulations reveal the monotonous

[a] Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France. E-mail: emmanuelle.bignon@univ-lorraine.fr
[b] Department of Biological, Chemical and Pharmaceutical Sciences. Universitá degli Studi di Palermo, via delle Scienze 90126 Palermo, Italy.
[c] Université de Paris, CNRS, ITODYS, F-75006, Paris, France, antonio.monari@u-paris.fr

† Electronic Supplementary Information (ESI) available: Computational details, main interactiion network and crucial protein/RNA distances. See DOI: 00.0000/00000000.

conformational behavior of the SL1 structure that exhibits flexibility mainly in its loop and to a lesser extent around the bulge. Interestingly, while rA27 is most frequently excluded from the helix, rC28 can remain in the major groove by interacting with the facing rA12 (ca. 60% of the simulation time). This slightly differs from the conformations recently reported by Bottaro et al.[13], yet we also observe transient conformations exhibiting an extruded rC28 and a helical rA27. These differences might be due to the presence of the 5' and 3' single-strand extremities in our model, which interestingly folds onto the hairpin structure and interacts with its backbone.

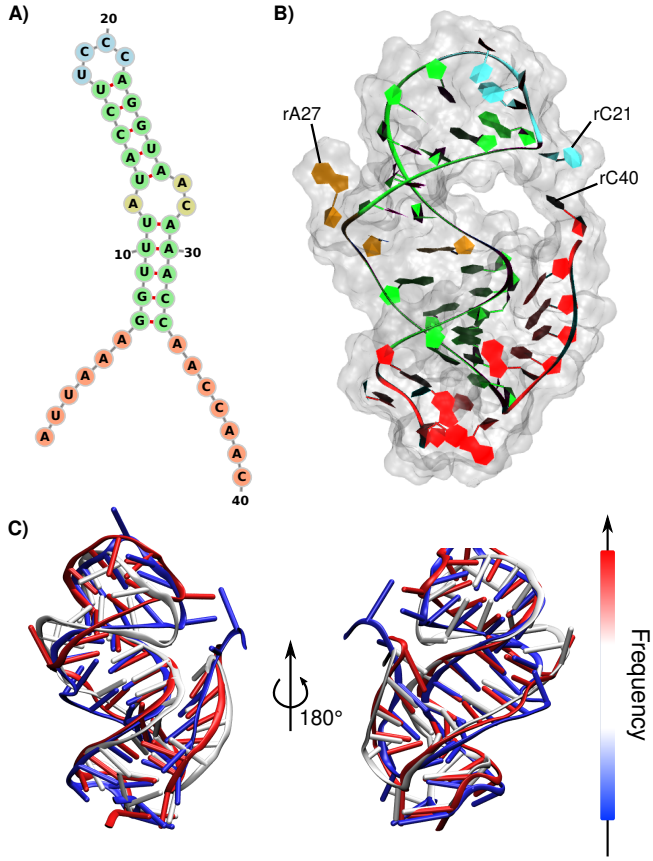

Fig. 1 A) Secondary structure of the 5'-UTR SL1 region of SARS-CoV-2 genome and B) a representative structure of the reconstructed 3D model. The loop, bulge, double-stranded, and single-stranded regions appear in blue, brown, green and red, respectively. C) Superimposed representative structures of the three major clusters extracted from the MD trajectory of the SL1 region.

Table 1 Total bend angle of the RNA for each system

system	average (°)	stdev (°)
SL1	104.3	34.6
OAS1-SL1	113.0	28.5
OAS1-RNA crystal	18.8	8.7

Additionally, the transient folding of the upper region of the hairpin onto the double-stranded region, observed in ca. 25% of the trajectory, allows to develop favorable interactions between rC21 and the rA12 backbone, which assist the extrusion of rC28. The rest of the time, the extruded rC21 is instead stabilized by π-stacking with 3'-rA39 and rC40. The presence of such extensive hydrogen bond network, involving the loop, the bulge, and the 5' single stranded regions imposes a strong bend to the SL structure as showed by Curves+[17] analysis (104.3 $\pm$ 34.6 °) - see Table 1. The Root Mean Square Deviation (RMSD)-based cluster analysis of the MD ensemble confirms that the loop is, indeed, the most flexible region of SL1, while the rest of the sequence exhibits a rather stable dynamic behavior - see Figure 1-C. As could be expected, the single-stranded regions also exhibit a pronounced flexibility, despite developing, as already stated, important interactions with the double-stranded region. Despite being interesting *per se* the characterization of the SL1 conformational space is also and foremost instrumental in providing crucial information for the study of its complexation with OAS1, as described in the following.

The most frequent conformation of SL1 was extracted from the MD ensemble via RMSD-based clustering and docked, using the HDock web server,[18] onto the crystal structure of OAS1, which has been retrieved from PDB 4IG8[19], to generate a starting model of the complex. The resulting 3D structure suggests that the 5' minor grrove of SL1 develops contacts with the OAS1 N-lobe while the second RNA minor groove, together with the loop region, are oriented towards and interact with the protein C lobe. On top of this initial structure we have performed equilibrium MD simulations using the NAMD code.[20,21] As detailed in the Supplementary Information the RNA-specific χ_{OL3}force field[22] has been used together with the AMBER FF14SB for the description of the protein.

Molecular dynamics simulations of the OAS1/SL1 complex, and of the reference crystal structure complexed with ideal double-stranded RNA,[19] highlights different interaction network in the two systems. Indeed, the strong bending of SL1, i.e. 113.0 $\pm$ 28.5 °, is even more pronounced upon complexation with OAS1 than for the isolated strand, see Table 1. This and the presence of the stem loop induce a very specific binding mode.

The two accessible minor groove regions of SL1 anchor the RNA to the OAS1 surface (see Figure 2), with hydrogen bonding between the RNA backbone/sugar and K42, R195, K199, and T203 in OAS1, similarly to what is observed for ds-RNA - see **Figures S1 and S2**. Interestingly, despite exhibiting the same contact regions between the RNA and the protein the interaction network is more extended in the reference ds-RNA/OAS1 system. The hydrogen bonds as described in the crystal structure[19] are persistent throughout the simulation, highlighting the non-specific interactions with ds-RNA backbone and sugar moieties - see Figure S3. This might be favored by the straighter structure of the reference double-stranded helix which exhibits an average bend of only 18.8 $\pm$ 8.7 °.

However, the same MD simulations allow to highlight the sequence dependency of the SL1 recognition by OAS1. Indeed, its specificity relies on the interactions with the RNA nucleobases. In the simulations of the reference ds-RNA system, such specific hydrogen bonds involve S56 backbone, Q158, K42, and T203, as

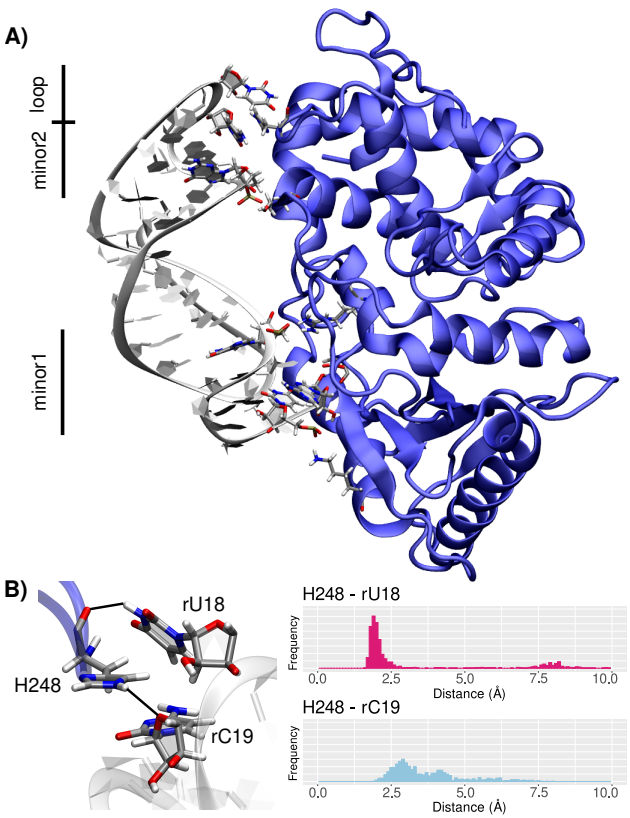

Fig. 2 A) Representative structure of the OAS1/SL1 complex exhibiting contact surfaces around the two minor groove and the RNA hairpin (loop). SL1 appears in white and OAS1 in blue, and the interacting amino and nucleic acids are displayed in licorice. B) Key-interactions in the loop region involving H248 and the RNA U18 and C19 (left), and their distribution along the MD simulation (right).

is initially seen in the crystal structure. On the other hand, the binding of SL1 nucleobases involves different residues, which are mainly spanning the extruded RNA bases and the loop region: T247 forms hydrogen bonds with rC19, and V58 and H248 backbone atoms interact with rA6 and rU18. The interaction of H248 with the RNA loop appears particularly important in SL1 binding to OAS1. Indeed, H248 is involved in a very persistent hydrogen bonds with rU18 through its backbone and with rC19 through its side chain - see Figure 2. Consequently, the OAS1/SL1 complex exhibits three important contact surfaces involving the loop and minor grooves of SL1 on the one hand, and both N and C lobes of OAS1. Interestingly, the bulge region of SL1 on the other hand is not involved in the interaction network with the protein. While the shift from rather unspecific backbone-driven to nucleobase-centered interaction may explain the affinity of OAS1 for the SL1 sequence, the question of the conservation of its sequence under the evolutionary pressure should be addressed. Miao *et al.*[7] have pointed out that the extruded bases on the bulge can be involved in base pairs in some variants, however their limited participation to the recognition should not hamper selectivity. As concerns the loop region, while the 5'-UTR and the stem-loop arrangement appear as fundamental for the viral replication[15], the SL1 region has nonetheless been recognized as a hotspot for point mutations. However the most common allele modifications involve the quartet rA34, rA35, rC36, and rC37[15], which seems less involved in OAS1 recognition.

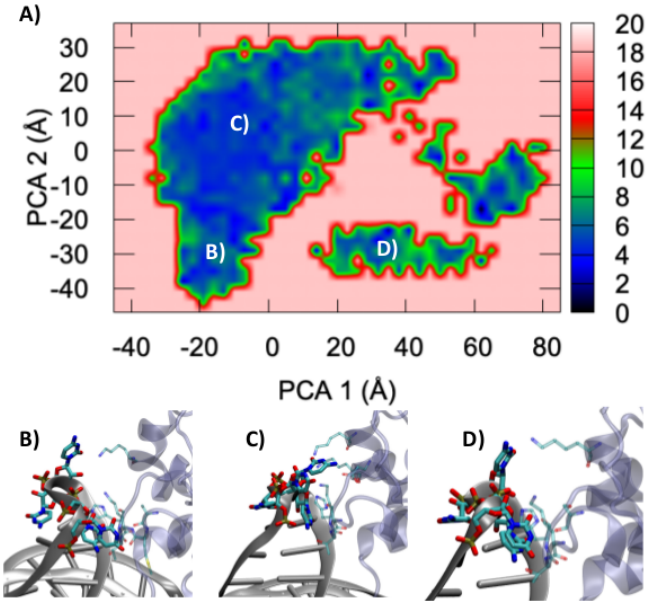

Fig. 3 A) Re-weighted free energy map obtained from the GAMD analysis as a function of the two main PCA modes. (B,C, and D) representative snapshots of the minimum regions, in which we have highlighted the loop area and the interacting residues, their location in the free energy map is also highlighted.

To ensure that our MD simulations have allowed a complete exploration of the conformational space of the OAS1/SL1 complex, and most notably described its most stable conformations, a Gaussian-Accelerated MD (GAMD) run,[23,24] in which an energetic repulsive bias is applied to all the dihedral angles of the protein and the nucleic acid, was performed to enhance the sampling of the phase space and avoid local minima traps. To take care of the effect of the bias, the re-weighted free energy map was obtained as a function of the projection of the MD trajectory on top of the two main RNA Principal Component Analysis (PCA) vectors, used as collective variables (Figure 3). As a matter of fact, the first two PCA vectors largely dominate the expansion and hence provide reasonable collective variables to describe the structural flexibility of our complex. From the analysis of the re-weighted free energy map one can evidence three main minimum basins. Interestingly, the principal basin represent a rather extended and flat energy surface which globally spans values of the first PCA vector (PCA1) comprised in the $-40/+40$ interval, showing only moderate energy penalties not exceeding 10 kcal/mol. The other two basins are, on the other hand, separated by higher barrier and are globally much less extended than the principal one. By analyzing representative structure belonging to the different basins, one can see that the main features already evidenced by equilibrium MD simulations are indeed preserved. In particular the main contact regions between the protein and the nucleic acid, as well as the main interaction network, are globally preserved, and still involve the minor groove region and the ex-

truded bases. On the same level the strong curvature of the RNA fragment is also maintained as well as the peculiar role of the uncoupled loop bases. Indeed, the persistence of the H248 interaction with the RNA bases located in the hairpin loop should be underlined. Likewise, the loop region experiences some flexibility, and is the area the more subjected to structural variation. In particular some rotations of the loop backbone dihedral are present and differentiate structures located in different basins, as well as in different regions of the large minimum. This backbone deformations are driven by the tendency to maximize hydrogen-bonds and polar interactions between the dangling nucleobases and the polar or charged OAS1 residues as shown in Figure 3. However, the high flexibility of the loop region and the global preservation of the main contact points confirm that the most important features of the interaction between OAS1 and SL1 as observed in our unbiased equilibrium MD simulations are actually conserved among the lowest energy conformations identified by enhanced sampling.

The response of the immune system to infections relies on the precise recognition of exogenous genetic or proteic material. The complex regulation of the innate immune system response to infections has been brought on the front line by the outbreak of the COVID-19 pandemics. Indeed, while an efficient immune response is needed to counteract the effects of SARS-CoV-2 infection, its deregulation has also been recognized as the cause of serious outcomes and morbidity. The recognition of exogenous viral RNA by the OAS1/RNAse L pathway is of particular importance in the innate response to SARS-CoV-2 infection. Indeed, it has been shown that variants of OAS1 inducing colocalization of the protein in the viral replication regions correlate with milder symptoms and better systemic response. Furthermore, a high *in vitro* affinity of OAS1 for the 5'-UTR of the viral genome, and in particular for the SL1 domain has been reported[1]. Yet, despite these important perspectives the structural and molecular bases driving this selectivity remain elusive and poorly characterized. In this contribution by using equilibrium and enhanced sampling MD simulations, we have provided a rationalization of the SL1 tertiary structure and of its specific interactions with the OAS1 RNA recognition region. In particular we have evidenced that the particular features of the SL1 moiety, and in particular its high bending, induces a slightly different interaction network compared to ds-RNA. More importantly, in addition to the contact with minor groove areas the interaction between the protein and the nucleic acid is also driven by the interaction with the dangling extrahelical bases. This is particularly true concerning the loop area, that exhibits some flexibility and is also strongly interacting with polar protein residues; the peculiar role of H248 in strongly anchoring the nucleic acid fragment has been evidenced. As a matter of fact, the interaction involving the hydrophobic nucleobase will certainly provide a higher degree of selectivity than salt bridges mainly involving the backbone phosphates, hence providing a rationale of the observed preference of OAS1 for the SL1 region. In the future we also plan to analyze the allosteric modulation of the OAS1 structure induced by the interaction with SL1. However, our results are already fundamentally important in providing an atomistic resolved vision on the reasons behind the recognition of SARS-CoV-2 genetic material by the innate immune systems, and hence can, in the long-term, help in designing therapeutic RNA strategy based on the stimulation of the immune system response by SL1 analogs.

The authors thank GENCI and Explor computing centers for computational resources. E.B. thanks the CNRS and French Ministry of Higher Education Research and Innovation (MESRI) for her postdoc fellowship under the GAVO program. AM thanks ANR and CGI for their financial support of this work through Labex SEAM ANR 11 LABX 086, ANR 11 IDEX 05 02. The support of the IdEx "Université Paris 2019" ANR-18-IDEX-0001 and of the Platform P3MB is gratefully acknowledged.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 S. Zhou, G. Butler-Laporte, T. Nakanishi, D. R. Morrison, J. Afilalo, M. Afilalo, L. Laurent, M. Pietzner, N. Kerrison, K. Zhao *et al.*, *Nature medicine*, 2021, **27**, 659–667.
2 A. Wickenhagen, E. Sugrue, S. Lytras, S. Kuchi, M. Noerenberg, M. L. Turnbull, C. Loney, V. Herder, J. Allan, I. Jarmson *et al.*, *Science*, 2021, **374**, eabj3624.
3 M. D'Antonio, J. P. Nguyen, T. D. Arthur, H. Matsui, A. D'Antonio-Chronowska, K. A. Frazer, C.-. H. G. Initiative *et al.*, *Cell Reports*, 2021, 110020.
4 J. Hu, J. Stojanovic, S. Yasamineh, P. Yasamineh, S. K. Karuppannan, M. J. H. Dowlath and H. Serati-Nouri, *Archives of Virology*, 2021, 1–24.
5 F. W. Soveg, J. Schwerk, N. S. Gokhale, K. Cerosaletti, J. R. Smith, E. Pairo-Castineira, A. M. Kell, A. Forero, S. A. Zaver, K. Esser-Nobis *et al.*, *Elife*, 2021, **10**, e710479.
6 S. L. Schwartz, E. N. Park, V. K. Vachon, S. Danzy, A. C. Lowen and G. L. Conn, *Nucleic Acids Research*, 2020, **48**, 7520–7531.
7 Z. Miao, A. Tidu, G. Eriani and F. Martin, *RNA biology*, 2021, **18**, 447–456.
8 C. Richter, K. F. Hohmann, S. Toews, D. Mathieu, N. Altincekic, J. K. Bains, O. Binas, B. Ceylan, E. Duchardt-Ferner, J. Ferner *et al.*, *Biomolecular NMR assignments*, 2021, **15**, 467–474.
9 L. Li, H. Kang, P. Liu, N. Makkinje, S. T. Williamson, J. L. Leibowitz and D. P. Giedroc, *Journal of Molecular Biology*, 2008, **377**, 790–803.
10 S. Zúñiga, I. Sola, S. Alonso and L. Enjuanes, *Journal of Virology*, 2004, **78**, 980–994.
11 L. Melidis, H. Hill, N. Coltman, S. Davies, K. Winczura, T. Chauhan, J. Craig, A. Garai, C. Hooper, R. Egan *et al.*, *Angewandte Chemie International Edition*, 2021.
12 N. Vankadari, N. N. Jeyasankar and W. J. Lopes, *The journal of physical chemistry letters*, 2020, **11**, 9659–9668.
13 S. Bottaro, G. Bussi and K. Lindorff-Larsen, *Journal of the American Chemical Society*, 2021.
14 A. Baldassarre, A. Paolini, S. P. Bruno, C. Felli, A. E. Tozzi and A. Masotti, *Epigenomics*, 2020, **12**, 1349–1361.
15 S. P. Ryder, B. R. Morgan, P. Coskun, K. Antkowiak and F. Massi, *Evolutionary Bioinformatics*, 2021, **17**, 11769343211014167.
16 M. Zuker, *Nucleic Acids Research*, 2003, **31**, 3406–3415.
17 C. Blanchet, M. Pasi, K. Zakrzewska and R. Lavery, *Nucleic Acids Research*, 2011, **39**, W68–W73.
18 Y. Yan, D. Zhang, P. Zhou, B. Li and S.-Y. Huang, *Nucleic Acids Research*, 2017, **45**, W365–W373.
19 J. Donovan, M. Dufner and A. Korennykh, *Proceedings of the National Academy of Sciences*, 2013, **110**, 1652–1657.
20 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781–1802.
21 J. C. Phillips, D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, *Journal of Chemical Physics*, 2020, **153**, 044130.
22 M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo and P. Jurečka, *Journal of Chemical Theory and Computation*, 2015, **11**, 5723–5736.
23 Y. Miao, V. A. Feher and J. A. McCammon, *Journal of Chemical Theory and Computation*, 2015, **11**, 3584–3595.
24 Y. T. Pang, Y. Miao, Y. Wang and J. A. McCammon, *Journal of Chemical Theory and Computation*, 2017, **13**, 9–19.