

Unraveling the existence of dynamic water channels in light-harvesting proteins: Alpha-C-phycoyanobilin in vitro

Hossam Elgabarty,^{abc} Peter Schmieder,^d and Daniel Sebastiani^{*abc}

Received Xth XXXXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXXXX 200X

DOI: 10.1039/b000000x

The three-dimensional conformation of proteins represents the basis for the great majority of functional investigations in biochemistry. Generally, this data is provided by X-ray diffraction techniques using crystalline samples. Here, we show that a considerably more accurate picture of the protein conformation in solution can be obtained by combining X-ray structures with first-principles calculations for the interpretation of two-dimensional solution-state NMR patterns. In the specific case of the bilin chromophore phycoyanobilin (PCB) in the binding pocket of the α subunit of C-phycoyanin (α -C-PC), we observe the formation of an unexpected water channel towards the PCB chromophore during molecular dynamics simulations. The resulting NMR pattern, computed from ab-initio methods in the newly obtained conformation, leads to a re-assignment of the experimental spectrum. The agreement with experimental data is considerably better than what we obtain using the unmodified X-ray structure. Generalizing our results, we elucidate the need to complement X-ray based crystal structures with experimental and computational solution-state spectroscopy, in order to reach a reliable representation of proteins in their native state. The latter often exhibits structural motifs like dynamical water channels, which are essential for the protein functionality but invisible in the crystal structure.

Introduction

X-ray crystallography and NMR spectroscopy are the main experimental techniques that provide protein structures at atomic resolution¹. Each of the two methods has its well-known advantages and technical limitations². Soon after the introduction of X-ray crystallography, it was accepted that it gives the correct picture of protein folding in spite of the non-physiological environment of a single crystal. However, some differences between solution-state and crystal structures of proteins are expected — and indeed known — to exist. For instance, in single crystals the protein molecules are densely packed, which may immobilise surface side-chains or even backbone segments that exhibit high mobility in the isolated protein³. As early as 1964, comparisons between NMR and X-ray structures of proteins were already being done to investigate differences between solution-state and crystal structures⁴. In the second half of the 1980s, NMR established itself as an alternative method for protein structure determination at a resolution comparable to that given by X-ray crystallogra-

phy. According to the statistics posted on the PDB website as of April 2012, the protein databank has a total of 74603 deposited protein structures. Approximately 10% of these were obtained using NMR, the absolute majority coming from X-ray crystallography.

Comparisons between protein structures obtained by both methods have always been highly interesting. Such comparisons necessarily require the availability of coordinates from both NMR and from X-ray crystallography for the same protein. This excludes proteins from different species, cases where one structure is for the free protein and the other for a complexed form, or where protein fragments do not have the same length³. Fortunately, with the exponential growth in the number of entries in the PDB, statistically significant comparisons are becoming more and more accessible. A large scale comparison on a set of 148 protein structure pairs revealed that in 73 proteins, the core heavy atoms are located at statistically different positions⁵. Several reasons have been suggested to explain these findings: Crystal packing effects, the presence of interactions (steric interactions, salt bridges, hydrogen bonds) in the crystalline state that are otherwise missing for the single protein, and the different refinement methods of both approaches.^{3,6,7} In another study on 78 proteins determined by both NMR and X-ray methods⁸, it was found that 18 proteins had obvious large-scale structural differences that seem to reflect a difference of crystal and solution structures. There was also a statistically-significant difference in the other 60 pairs

^a Physics Department, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany. Fax: ++49 (0)30 838 56046; Tel: ++49 (0)30 838 56149; E-mail: daniel.sebastiani@fu-berlin.de

^b Dahlem Center for Complex Quantum Systems (DCCQS)

^c The Max Planck Institute for Polymer Research (MPIP), Ackermannweg 10, 55128 Mainz, Germany.

^d Leibniz-Institut für Molekulare Pharmakologie (FMP), Robert Rössle-Str. 10, 13125 Berlin Germany.

of structures in inter-residue contacts, which was mainly attributed to a difference in mathematical treatment of experimental results. In addition to such static structural differences, deviations in the dynamics of NMR and X-ray protein structures are also known to exist. As one might expect, the large-amplitude motions sampled in solution are restricted in protein crystals.

In a study by Snyder et al.¹ using data on 159 proteins, it was shown that the majority of proteins with high quality NMR data suitable for 3D structure determination, do not rapidly and readily yield diffraction quality crystals, and vice-versa. It was concluded that “X-ray crystallography and NMR often provide complementary sources of structural data and that both methods are required in order to optimize success for as many targets as possible in large-scale structural proteomics efforts”.

Here we present a study of the alpha-subunit of C-Phycocyanin, with focus on its bilin chromophore, phycocyanobilin (PCB), and its binding pocket. We exploit the high sensitivity of ab initio NMR calculations as an extremely sensitive local probe at the sub-Angstrom range, to reveal structural differences between the crystal structure and the solution state structure. More generally, we propose an ab initio approach that can accurately account for local geometric differences between the X-ray structure of a protein and its solution structure, requiring only a minimal set of experimental NMR measurements.

Bilins are linear tetrapyrrolic pigments with a conjugated bond structure. Traditionally, the four rings are referred to as ring **A**, ring **B**, ring **C**, and ring **D**. Bilins are ubiquitous in nature, being found in animals, plants, algae, and bacteria. In photosynthetic organisms bilin chromophores serve as prosthetic groups for the photoreceptor *biliproteins*, where their conjugated bond structure confers the desired wavelength sensitivity to the apoprotein. They are typically attached to the apoprotein via a thioether link to a conserved cysteine residue⁹.

Bilins provide a clear example of how far the local environment can influence both the dynamical and the photochemical properties of chromophores. The most straightforward case is the dependence of the absorption spectrum of the free chromophore in solution on its protonation state¹⁰. Moving on from free bilins in solution to biliproteins, we find that the same bilin chromophore in different biliproteins exhibits different absorption wavelengths. Obviously, the local protein environment at the chromophore binding pocket can significantly influence the electronic structure of the embedded chromophores, optimizing them for their light-sensing or light-harvesting functions^{11–13}. A recent manifestation of the strong potential of these protein-bilin interactions to modify the behavior of the embedded chromophore was obtained using DFT methods¹⁴. The geometric relaxation of free un-

bound bilins following excitation has been found to be dominated by torsional motion between rings **A** and **B** (see Fig. 1), and irrespective of the initial configuration. This is in clear contrast to the situation when the chromophore is attached to a phytochrome where the bulk of experimental evidence strongly suggests a photo-induced torsional motion between rings **C** and **D**^{15–22}.

Two major classes of biliproteins are phytochromes and phycobiliproteins. Although these two classes share the same type of prosthetic groups, they have totally different functions and photochemical behavior. Phytochromes are sensory red/far-red photoreceptors that exist in two states, a red-absorbing state (P_r) and a far-red-absorbing one (P_{fr}). When exposed to light of the suitable wavelength the chromophore undergoes a light-dependent conformational change which propagates through the phytochrome protein to a response domain²³. The cyanobacterial phycobilisomes on the other hand act as light-harvesting antennae for photosystems I and II. The chromophore here does not undergo a light-dependent conformational change, but the photon energy is rather transferred efficiently down a cascade of multiple chromophores until it reaches chlorophyll, where photosynthesis takes place²⁴. The highly efficient light-harvesting of these proteins has been a subject of extensive research. Recent multidimensional nonlinear spectroscopic experiments provide compelling evidence that the excitation energy transfer in such light-harvesting systems involves relatively long-lived coherent quantum dynamics^{25,26}.

C-Phycocyanin (C-PC) is one of the major biliproteins in phycobilisomes, and one of the earliest to be studied²⁷. Besides its role as a photoreceptor, C-PC has attracted much biomedical attention with the numerous reports of its therapeutic and nutritional values²⁸. C-PC exists in solution as a complex of trimers ($\alpha_3\beta_3$), hexamers ($\alpha_6\beta_6$), and other oligomers, where each α -subunit has one bilin chromophore known as phycocyanobilin (PCB) (Fig. 1 left) and each β -subunit has two PCBs²⁹. Of significance to our work here is the observation that the alpha-subunit does not form aggregates in solution at the typical concentrations used for NMR measurements³⁰. This triggers the question of whether the isolated alpha-subunit in solution retains the same structure as found in X-ray crystal structures or not.

Like other bilins, PCB is composed of four pyrrole rings linearly attached together with ring **A** being the one attached to the apoprotein, and ring **D** is the furthest. Since we will be frequently referring to the pyrrole nitrogen atoms and the hydrogens bonded to them, we will refer to them in the rest of this article as **NA**, **NB**, ..., **HC** and **HD**. The conformation of a bilin chromophore is conventionally described with respect to the three double (*cis/trans*) and three single (*syn/anti*) bonds comprising the methine bridges between the pyrrole rings (Fig. 1 right).

The structure and function of C-PC and PCB have been investigated by numerous methods, both experimental and theoretical. Through these studies, a fairly reliable picture of the structure of C-PC has been formulated and several high-resolution X-ray crystal

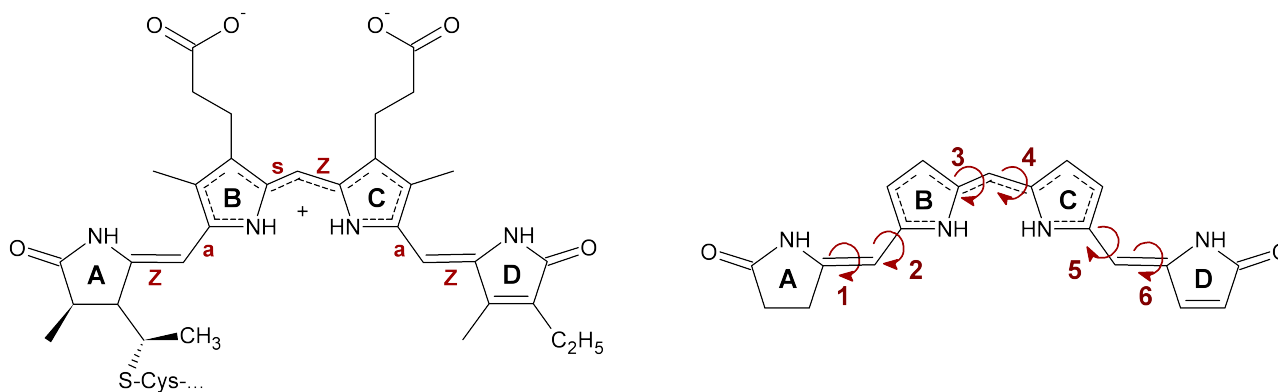


Fig. 1 Left: The structure of phycocyanobilin (PCB). Ring A is attached via a thioether link to a conserved cysteine residue in the apoprotein. This two-dimensional structure corresponds to a *ZZZasa* conformation. Right: The six bonds that are traditionally used to define the conformation of a bilin. The three double bonds are denoted by *Z/E* while the three single bonds by *syn/anti*.

structures are now available (e.g. PDB entries: 3KVS, 3L0F, 3O18 and 3O2C)^{31–33}. A two-dimensional projection of PCB in the binding pocket is depicted in Fig. 2. It can be seen that PCB adopts a *ZZZasa* conformation. PCB is tightly held in place via its numerous interactions with the binding site. Ring A is the only one buried in the binding pocket without any access to solvent. Its carbonyl oxygen atom possibly makes a hydrogen bond with the backbone amide of ALA75 (distance 2.10 Å). The closest hydrogen bond acceptor to HA is the backbone carbonyl of ASN73, which is 2.65 Å away and makes an angle of 133 degree with the NA-HA bond. This length/angle combination is outside the usual range for a medium-strength hydrogen bond³⁴. Rings B, C, and D are all positioned such that they have one edge exposed to the solvent. The two nitrogen atoms of ring B and C each donates a hydrogen bond to the same carboxylate oxygen atom in the ASP87 side chain. The two propionate side chains of rings B and C are oriented outside the binding site and towards the solvent. One of the carboxylate oxygens of ring B propionate forms a salt bridge with the positive ARG79 side chain and the other oxygen is solvated by water, while ring C propionate is twisted back so that both oxygens are involved in salt bridges with the positive LYS83 and ARG86 side chains. Ring D seems to be the most solvent-accessible one, both the carbonyl oxygen and HD are fully exposed to the solvent.

NMR has also been used to investigate the mobility of the chromophore in the binding site³⁵, solution state ¹⁵N NMR of the α subunit of C-PC was used to deduce the protonation state of PCB in the binding pocket and to infer a structural picture of PCB in the cyanobacterial phytochrome Cph1³⁶, QM/MM calculation of Raman spectra were performed and showed good agreement with experimental spectra³⁷. Through the accumulated results of all these studies one can say with confidence that the conformation of PCB in α -C-PC is indeed *ZZZasa*. NMR also provides convincing evidence that all the four pyrrole nitrogens of PCB are protonated in C-PC, this seems to be also true for other biliproteins^{16,17,36}. The significance of these results is not only restricted to C-PC and its chromophore, PCB. Since the availability of the first X-ray structures of C-PC they have served as models for understanding the structure and function of

other biliproteins, in particular phytochromes as their tertiary structures have begun to come out only recently.

In a relatively recent study, the ¹⁵N NMR spectrum of α -C-PC from *Mastigocladus laminosus* has been compared to that of the cyanobacterial phytochrome Cph1, providing insight into the structure and dynamics of the binding pocket of the latter³⁶. In that study, the ¹H,¹⁵N HMQC NMR spectrum of labeled α -C-PC revealed five coherences with a nitrogen chemical more than 130 ppm (Fig. 3). Of these five signals, two showed an interaction in the NOESY spectrum and were thus identified as HB and HC due to their proximity. In the original published assignment a peak showing an interaction with two protons in the 7.0-7.5 ppm range, and was identified as HA due to its proximity to the back bone amine protons of Ala75 and Tyr74. HD did not show any interactions, possibly due to its location at the edge of the binding pocket facing the bulk solvent, and the correlation at 149.4/9.5 ppm was assigned to it. Such assignment seemed to be the most plausible initially, however, based on more recent NMR work on isolated PCB in HMPT³⁸, and from solid-state NMR of cyanobacterial phytochrome (Cph1)¹⁷, we tend to believe that the initial assignment of ring D, and possibly ring A, might not be correct. (we need to put here a figure from the supp. info from Schmieder's paper) In both of these publications, ring D gave a resonance close to 130 ppm, and we believe that PCB ring D in α -C-PC is actually the resonance showing at 132.4/9.3, close to the protein backbone resonances. If this is the case, then ring A might actually be the resonance that was initially assigned to ring D, and then the resonance at 168/12.4 ppm was one of the protein histidines. However, one has to be cautious when comparing the NMR spectrum of PCB in α -C-PC to that of the free PCB, or even to PCB bound to Cph1, where PCB poses in a *ZZZssa* geometry, surrounded by a different environment.

With the current availability of high-resolution structures of biliproteins, it is now possible utilize molecular dynamics simulations based on these structures to study the mobility and dynamics of the chromophore and the atomistic details of its interactions with the apoprotein. This is particularly important in order to facilitate an explanation of the various differences in the behavior of different

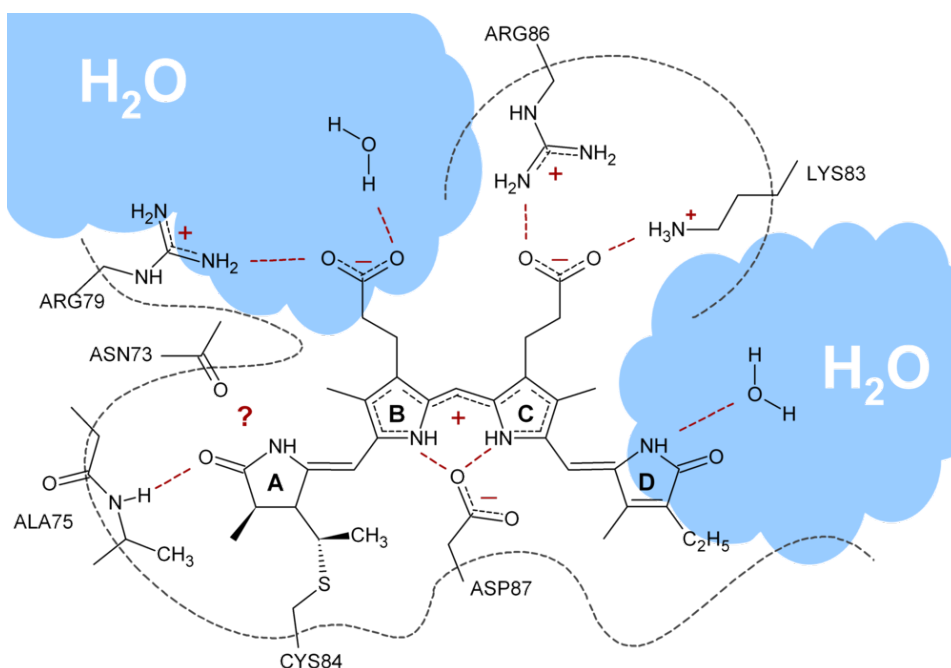


Fig. 2 A two-dimensional projection of PCB in the α -C-PC binding pocket showing the various interactions with the binding site.

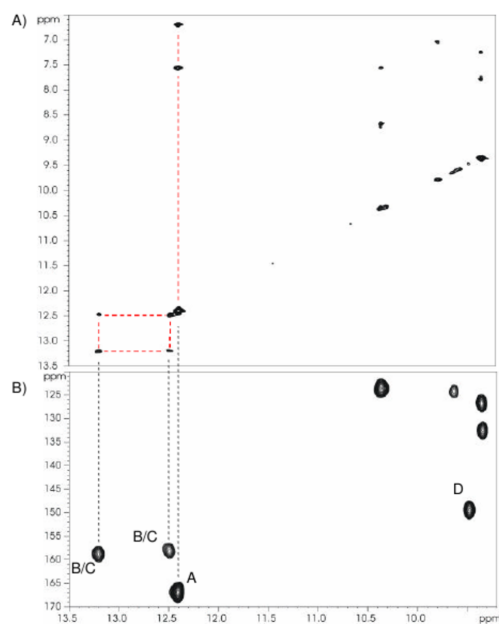


Fig. 3 The initial assignment of the four nitrogen bound protons of PCB in α -C-PC from *Mastigocladus laminosus*.³⁶ A) NOESY spectrum. The two protons that show an interaction with each other are HB and HC. B) Region from the ^1H , ^{15}N HMQC spectrum.

classes of phytochromes that cannot be explained based on the static X-ray structural pictures alone. In this particular case of PCB bound to α -C-PC, it is also tempting to exploit the availability of both an X-ray structure of the system and the NMR spectrum. One is tempted to utilize the sensitivity of NMR shifts to examine any possible differences between ab initio H and N chemical shifts computed from the crystal structure, and the solution state HMQC NMR spectrum.

Here, we present first-principles QM/MM MD simulations and theoretical calculations of NMR chemical shifts of PCB in the α -C-PC binding pocket. One goal of this work is to gain insight into the (mainly) non-bonded interactions between PCB and its binding pocket, and how these influence the conformation and dynamical behavior of PCB. Experimental NMR chemical shifts of PCB bound to proteins with a resolved X-ray structure have been used in attempts to deduce the geometry of PCB and other related chromophores in proteins with unavailable tertiary structures. It is an interesting question to see whether theoretical calculations of NMR chemical shifts combined with the insight obtained from MD simulations can help push forward these attempts, and to draw conclusions on other related chromophores. Another goal is to test the convergence of the computed chemical shifts in such a complicated system with regards to the embedding scheme, i.e. how the computed values respond to the level of sophistication in treating the surroundings of the chromophore.

Computational details

Structure preparation and classical MD simulations

As a starting point we used the X-ray structure of the C-Phycocyanin trimer from *Mastigocladus laminosus*^{39,40}. The α -subunit was extracted from the structure, missing hydrogens were added and the structure was solvated in a water box with dimensions 70 x 90 x 60 Angstroms. Based on the experimental NMR evidence detailed in the introduction, all the four pyrrole rings of PCB were protonated, while the propionate side chains were left ionized so that PCB molecule had a total charge of -1. The system was equilibrated for 50 ns at 300 °K with classical MD using the CHARMM22 forcefield^{41,42} in the NAMD program⁴³. CHARMM22-compatible force field parameters for PCB are already available and were used here.⁴⁴

Ab initio QM/MM MD simulations

QM/MM MD simulations were performed in CP2K⁴⁵⁻⁴⁷. The QM region consisted of the PCB chromophore, the attached CYS84 side chain, and the ASP87 side chain. The rest of the system was treated with the CHARMM22 forcefield. The QM/MM bond interfaces in CYS84 and ASP87 were modeled using an optimized capping potential^{48,49}. The size of the QM box was set to 30 x 30 x 30 Angstroms. The QM part was treated with the BLYP functional with the Grimme D2 correction⁵⁰ using the GPW method and a planewave density cut-off of 280 Ry and a triple zeta gaussian basis set. 20 ps of MD were run for equilibration under massive thermostating, then finally a production run of 20 ps was performed using a Nose-Hoover thermostat at 330 °K. Each of the QM and the MM subsystems was coupled to a separate thermostat. For NMR calculations, 100 snapshots were extracted at regular intervals from the production phase of the simulation.

Nuclear shieldings

For each of the snapshots extracted from the MD trajectory, all-electron isotropic magnetic shieldings were computed for the four pyrrole nitrogens and hydrogens. CP2K was used to compute the shieldings in a QM/MM setup. In all cases the QM region encompassed PCB and all hydrogen-bonding residues to any of the pyrrole hydrogens. Additional QM/MM capping atoms were introduced to encompass all the relevant atoms in the QM region (Fig. 4). The all-electron GAPW method⁵¹ was used with the pcS-2 basis set (pcS-3 on the pyrrole nitrogens and hydrogens)⁵². Proton chemical shifts were referenced to TMS at the same level of theory, while nitrogen chemical shifts were referenced to ammonia and then shifted slightly so that the center of the computed shifts matched that of the experimental ones.

Results and Discussion

Dynamics of PCB in the binding pocket

Fig. 5 depicts the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) of PCB non-hydrogen atom over

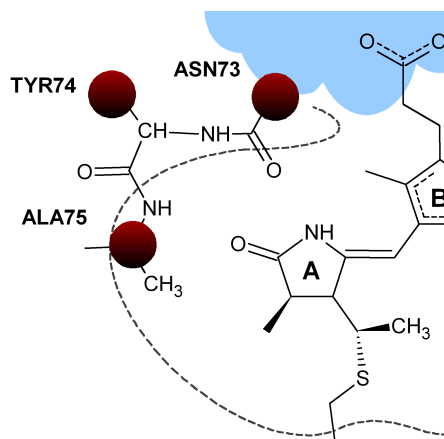


Fig. 4 QM/MM partitioning used for the NMR calculations. The balls correspond to link atoms (capping atoms) which saturate the QM region⁴⁹. Here specifically the QM region is the backbone segment from residue 73 to residue 75 excluding the side chains.

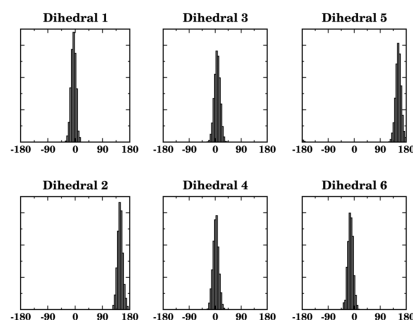


Fig. 6 Histogram of the six dihedral angles in PCB from 20 ps MD trajectory (See Fig. 1 for the definition of the angles). The abscissa is in arbitrary units.

the QM/MM trajectory. The positions of PCB ring atoms show little fluctuations, with the exception of ring D and its ethyl side chain. Fig. 6 shows the statistical distribution of the six dihedral angles in PCB defined as in Fig. 1. The distributions indicate that the single/double bond pattern depicted in Fig. 1 is correct, with the bond resonance in rings B and C giving a partial double bond character to the two central bonds connecting them.

Overall, PCB shows very little mobility in its binding pocket. The lack of mobility of ring A can be attributed to its attachment to the protein, while that of rings B and C can be attributed to them being firmly sandwiched between the hydrogen bonds to ASP87 on one side, and the ion pairing between their propionate groups and ARG79, LYS83, and ARG86 on the other side (Fig. 2).

Another interesting point is the protonation state of rings B and C. Fig. 7 shows that both HB and HC protons remain firmly attached to their nitrogens, in agreement with NMR spectra which show that the protons remain attached to the pyrrole nitrogens even on a much larger time scale. Ab initio thermochemical predictions provide a

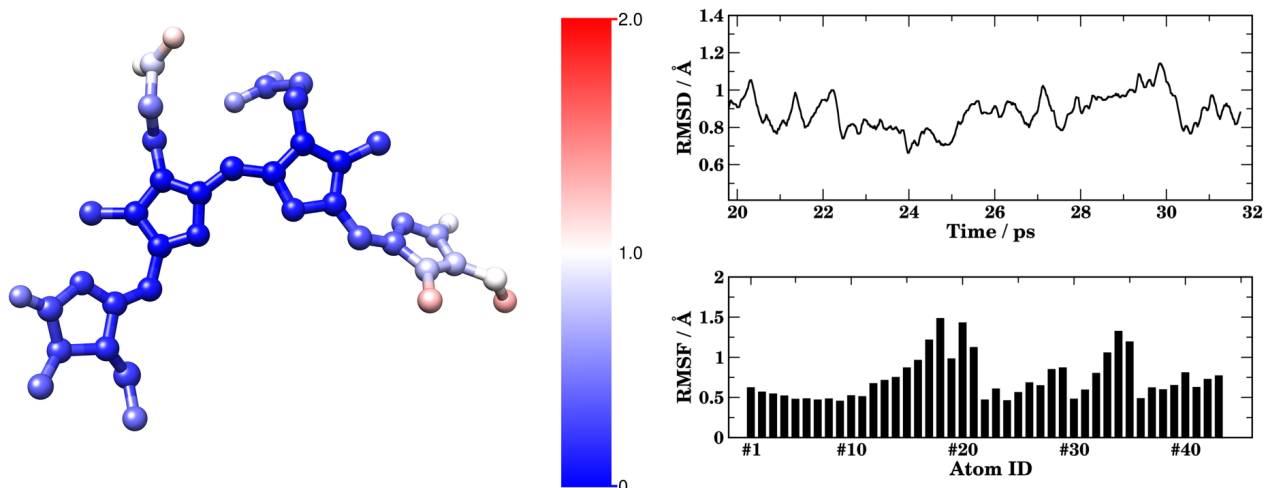


Fig. 5 Mobility of PCB during the MD in terms of the RMSD and RMSF. The RMSF is color-coded in the PCB structure. Except for ring **D**, ring atoms show very little mobility.

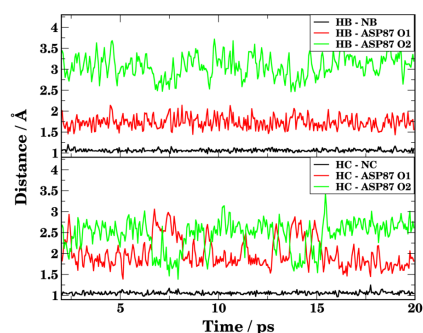


Fig. 7 Top: Distances between **NB** and: (1) **HB**, (2) the two carboxylate oxygens of the **ASP87** side chain. Same distances for **NC**.

similar picture, where the activation energy to transfer one proton from the **ASP87** to **PCB** was calculated to be 0.76 kcal/mol, and the reverse barrier was found to be 5.57 kcal/mol at the level of B3LYP/6-31+G*//HF/6-31G⁵³.

NMR signatures of hydrogen bonding

Fig. 8 shows the NMR chemical shifts of **PCB** pyrrole nitrogens and hydrogens computed from first principles as an ensemble average over the configurations delivered by the MD trajectory. The corresponding experimental values are also shown for comparison. The points corresponding to the experimental spectrum are labeled according to the initial assignment³⁶, while the point labeled as \tilde{D} is what we believe to be the correct assignment of ring **D**. Overall, rings **B** and **C** show very good agreement, and interestingly, the ab initio value for ring **D** confirms that \tilde{D} is indeed the corresponding exper-

imental signal. The predicted nitrogen chemical shift for ring **A** is very close to the experimental signal for the point assigned as **A**, while it is significantly off for the hydrogen chemical shift. Given that nitrogen chemical shifts for all the three other **PCB** rings are in very good agreement, we tend to believe that the initial assignment for ring **A** is correct.

Further insight into the correlation between binding pocket geometry and the NMR signature is obtained by examining the correlations depicted in **Fig. 9**. The plots show the calculated dependence of instantaneous proton chemical shifts on the length of the corresponding hydrogen bond. In the region of strong hydrogen bonding ($d_{H-D} < 2.2$) the relation is approximately linear, followed by an asymptotic decay for longer distances towards the chemical shift of the non-hydrogen-bonded situation. From the plot corresponding to **HA**, we can see that its chemical shift corresponds to a hydrogen bond length averaged around 1.8 Å. Looking at the slopes of the linear regressions considering only lengths up to 2.0 Å, we find that all the plots exhibit a proportionality in the range -9.5 to -10.5 ppm/Å, except for ring **A** which a lower slope of -8.0 ppm/Å. These large values for the slope are the very reason why NMR is such a good probe of local geometries: We have a change in the observed ¹H-chemical shift of 1 ppm (a value easily within the accuracy of both experiment and ab initio methods) per 0.1 Å change in the hydrogen bond length. One can even go further and use such a plot as a “calibration curve” to correlate the observed chemical shifts to local geometries at the sub-Angstrom range, a range that is inaccessible for most other spectroscopic techniques.

Going back to our trajectories, we depict in **Fig. 10** the time-evolution of the distances between ring **A** and the nearest possible hydrogen bonding groups, from both the **QM/MM** and the **CHARMM22** MD trajectory. We can see that these distances are too long to influence the chemical shift of **HA**, they are within the asymptotic part in **Fig. 9**, not to mention to explain the strong shift seen in the experimental NMR spectrum. These results point in the

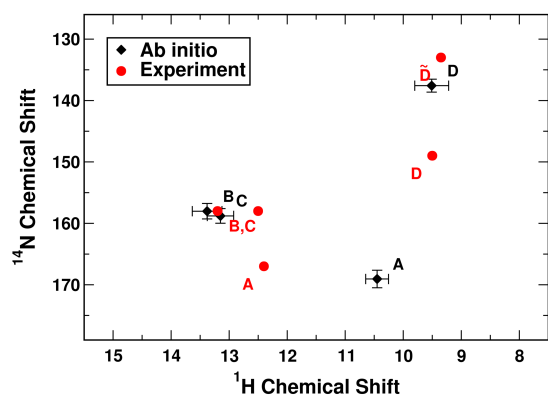


Fig. 8 Experimental vs. MD-averaged ab initio NMR chemical shifts of the pyrrole nitrogens and hydrogens in PCB. Here, the overall geometry of PCB and its binding pocket correspond closely to that depicted in Fig. 2. The point labelled as **D** is what we believe to be the actual experimental NMR signal of ring **D**.

same direction as our arguments in the introduction, the X-ray structure and the solution-state NMR spectrum do not match for ring **A**.

Our first attempt to reconcile the X-ray structure and the solution-state NMR results was running another CHARMM22 MD simulation, this time constraining the distance between **HA** and the backbone oxygen of ASN73 to 1.8 Å. In this way, a strong hydrogen bond is enforced for **HA**. After equilibration, the system was left to evolve for 40 ns under this constraint before releasing it and allowing the system to evolve freely for 60 ns. As Fig. 11 shows, the protein backbone starts gradually recoiling away from **HA** once the constrain is gone, and within 10 ns, all memory of the constrain is lost. In conclusion, this attempt did not succeed to “convince” the chromophore to attain a hydrogen bonding configuration which would presumably yield a computed NMR chemical shift pattern compatible with experiment.

Another interesting possibility arises by careful examination of our initial, unconstrained CHARMM22 trajectory. Fig. 12 depicts two representative configurations of the binding pocket where water wires are forming between PCB, ALA75, ASN73, and bulk water. The initial event for formation of such water wires is the opening up of the “binding pocket gate” near ring **A**: A movement of the side chain ASN73 away from PHE122 backbone, exchanging the lost hydrogen bond by another stable one to the propionate side chain of ring **B**. Such an arrangement can be seen in the right side of Fig. 12. Once the gate is open bulk water starts pouring inside the pocket, resulting in a water wire as depicted in Fig. 12. The backbone amino group of ALA75 can be seen switching between two situations: in one it is hydrogen-bonded to the oxygen atom of ring **A** (Fig. 12 left, anti-conformation relative to ASN73), and in the other it is forming a hydrogen bond to LEU66O (Fig. 12 right, clinal-conformation), which is part of an alpha-helix on the solvent-exposed protein surface. Indeed, such configurations can also be seen forming in the last few picoseconds in the ab initio MD trajectory. In fact, if these con-

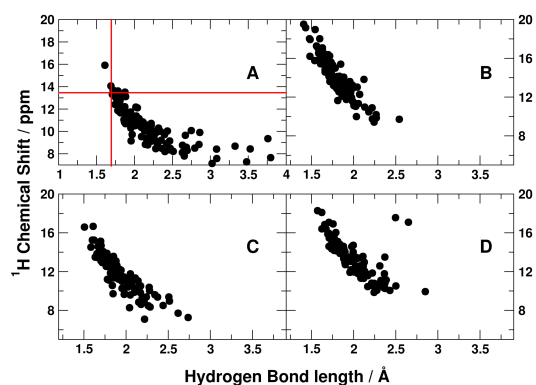


Fig. 9 Correlation between hydrogen bond length and the chemical shift of the pyrrole hydrogens

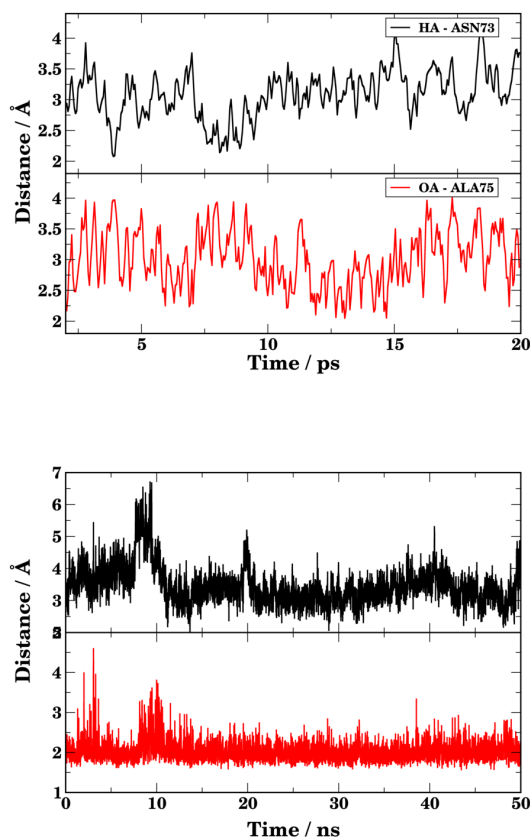


Fig. 10 Top: Time evolution of the hydrogen bond length between **HA** to ASN73 backbone oxygen (top), and between ALA75 backbone hydrogen to ring **A** carbonyl oxygen (bottom). Bottom: Same distances taken from the CHARMM22 trajectory.

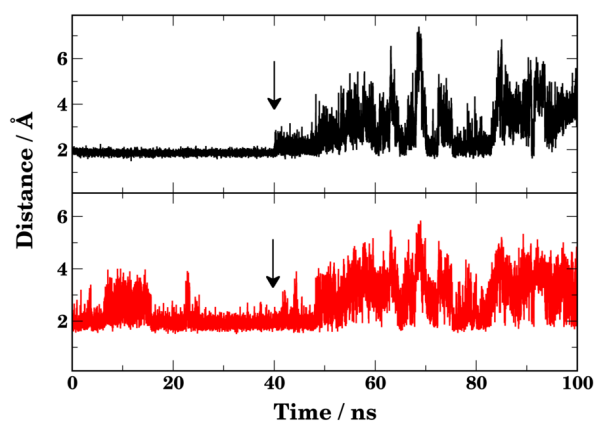


Fig. 11 Time evolution of the HA-ASN73 (Top), and ring A oxygen-ALA75 distances (Bottom). The arrows mark the time when the distance constraint was removed.

figurations are excluded from the NMR sampling, HA chemical shift further moves by 0.5 ppm upfield.

Based on this evidence, we started another QM/MM trajectory from such a configuration with water wires, using the same computational setup as before for both the MD and the NMR calculations. The MD reveals that once a water molecule is trapped between HA and the ASN73 backbone, it forms a very stable bridge, with an average distance of 1.85 Å between the water oxygen and HA. In fact, such a configuration agrees with the experimental NMR signature, as it represents a stable hydrogen bond to ring A at the required distance. Fig. 13 shows the MD-averaged NMR chemical shifts (20 ps) from the new setup compared to experimental results. The agreement between computed and experimental NMR signatures is far better in this case, confirming that the inclusion of the water wire at ring A yields an NMR chemical shift pattern that matches the experimental counterpart very well.

As for the correct assignment of the originally mislabeled NMR resonances, we believe that it might have originated from a protein histidine. However at the present state, we do not have sufficient experimental/computational evidence to claim that this is a certain assignment.

Conclusions

We have presented an analysis of the accessibility of the bilin chromophore embedded in the α -subunit of C-phycoyanin with respect to water molecules from within the protein. We have used first-principles molecular dynamics simulations and ab-initio calculations of NMR chemical shift patterns to interpret the corresponding experimental data. The comparison of computational and experimental data leads to considerable structural differences between the solution-state and the crystalline conformations, in particular in view of the microsolvation of ring A of the bilin chromophore. The loop in the PCB

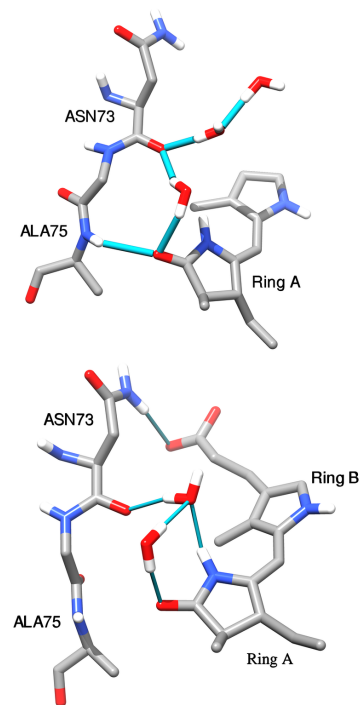


Fig. 12 Examples of the water wires that form inside the PCB binding pocket, bridging ring A to ASN73.

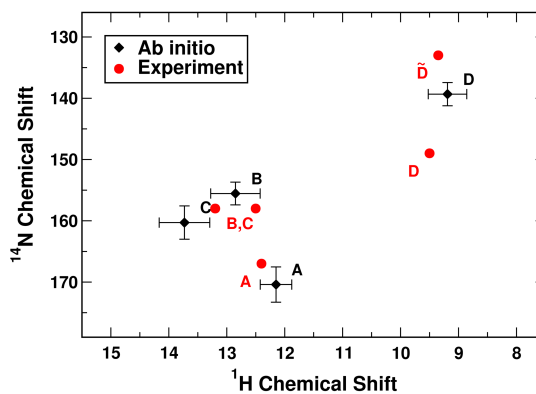


Fig. 13 MD-averaged NMR chemical shifts taken from the QM/MM trajectory in which a water wire bridges ring A to ASN73 and connects to bulk water (Fig. 12).

binding site, which is packed against the crystal packing surface, has more freedom in solution which leads to its movement away from PCB. This allows one water molecule to pour inside in the binding pocket, bridging ring A to ASN73 backbone, and eventually forming a stable water wire that connects to bulk solvent.

Our approach of computing ensemble averages of NMR chemical shift patterns from first principles molecular dynamics simulations provides a working method to reconcile X-ray crystal structures and solution state structures to reconcile X-ray crystal structures and solution state structures using a minimal set of experimental NMR data. The speed with which NMR chemical shifts can be computed on modern computers makes the method quite affordable, even with the extensive sampling required to converge the NMR shifts towards the experimental results. Another possible application of this approach is for membrane proteins with poorly resolved structures. If one is particularly interested in one region, then our method provides a microscopic probe that can look at distances at the sub-angstrom range provided that experimental NMR/EPR signatures are available. The argument applies equally well to protein structures that cannot be investigated by NMR (large proteins). One has to keep in mind that the choice of the size of the QM region and the capping method in the MD/NMR calculations is a very critical choice. The quality of the capping method becomes particularly critical when the capping atom is close to an atom where NMR chemical shift is to be computed.

We view this work as a continuation of the efforts that try to utilize ab initio QM methods in structural biology, one particular example is the use of ab initio methods in optimizing X-ray data.⁵⁴ We go here one step further by combining MD and NMR calculations to provide results directly comparable to a highly sensitive experimental observable. Our calculations show that a fully consistent first-principle treatment, from molecular dynamics simulations to ensemble averages of spectroscopic observables, can elucidate subtle differences between crystal structures and solution structures of proteins. In our case, the local mobility of a particular building block of the chromophore is considerably increased under solvated conditions at ambient temperatures, which is connected to an opening of the binding pocket towards a neighboring water channel.

Acknowledgements

Financial support is gratefully acknowledged from the DFG under grant Se 1008/6 and from the NanoScale initiative of the FU Berlin. Supercomputer time was provided by the Julich supercomputing center under grant 4781. Molecular graphics images were produced using VMD and the UCSF-Chimera package.

References

- 1 D. a. Snyder, Y. Chen, N. G. Denissova, T. Acton, J. M. Aramini, M. Ciano, R. Karlin, J. Liu, P. Manor, P. a. Rajan, P. Rossi, G. V. T. Swapna, R. Xiao, B. Rost, J. Hunt and G. T. Montelione, *J. Am. Chem. Soc.*, 2005, **127**, 16505–11.
- 2 J. Kirchmair, P. Markt, S. Distinto, D. Schuster, G. M. Spitzer, K. R. Liedl, T. Langer and G. Wolber, *J. Med. Chem.*, 2008, **51**, 7021–40.
- 3 M. Billeter, *Q. Rev. Biophys.*, 1992, **25**, 325–377.
- 4 M. Praissman and J. Rupley, *J. Am. Chem. Soc.*, 1964, **86**, 3584–3585.
- 5 M. Andrec, D. A. Snyder, Z. Zhou, J. Young, G. T. Montelione and R. M. Levy, *Proteins*, 2007, **69**, 449–65.
- 6 I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns and J. Yuan, *J. Am. Chem. Soc.*, 2009, **131**, 5134–44.
- 7 P. R. Kuser, L. Franzoni, E. Ferrari, A. Spisni and I. Polikarpov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2001, **57**, 1863–1869.
- 8 L.-W. Yang, E. Eyal, C. Chennubhotla, J. Jee, A. M. Gronenborn and I. Bahar, *Structure*, 2007, **15**, 741–9.
- 9 A. T. Ulijasz and R. D. Vierstra, *Curr. Opin. Plant Biol.*, 2011, **14**, 498–506.
- 10 W. Rüdiger and H. Scheer, *Photomorphogenesis*, Springer: Berlin, 1983, vol. 16, ch. 7, pp. 119–151.
- 11 S. E. Braslavsky, A. R. Holzwarth and K. Schaffner, *Angew. Chem., Int. Ed.*, 1983, **22**, 656–674.
- 12 H. Scheer, *Angew. Chem., Int. Ed.*, 1981, **20**, 241–261.
- 13 J.-M. Tu, M. Zhou, R. Haessner, M. Plöschner, L. Eichacker, H. Scheer and K.-H. Zhao, *J. Am. Chem. Soc.*, 2009, **131**, 5399–401.
- 14 A. Strambi and B. Durbeej, *Photochem. Photobiol. Sci.*, 2011, **10**, 569–79.
- 15 W. Rüdiger, F. Thümmel, E. Cmiel and S. Schneider, *PNAS*, 1983, **80**, 6244.
- 16 H. M. Strauss, J. Hughes and P. Schmieder, *Biochemistry*, 2005, **44**, 8244–50.
- 17 T. Rohmer, C. Lang, J. Hughes, L.-o. Essen, W. Gärtner and J. Matysik, *PNAS*, 2008, **105**, 15229–34.
- 18 M. A. Mroginski, D. H. Murgida and P. Hildebrandt, *Acc. Chem. Res.*, 2007, **40**, 258–266.
- 19 M. A. Mroginski, D. von Stetten, F. V. Escobar, H. M. Strauss, S. Kaminski, P. Scheerer, M. Günther, D. H. Murgida, P. Schmieder, C. Bongards, W. Gärtner, J. Mailliet, J. Hughes, L.-O. Essen and P. Hildebrandt, *Biophys. J.*, 2009, **96**, 4153–63.
- 20 K. Inomata, M. a. S. Hammam, H. Kinoshita, Y. Murata, H. Khawn, S. Noack, N. Michael and T. Lamparter, *J. Biol. Chem.*, 2005, **280**, 24491–7.
- 21 K. Inomata, H. Khawn, L.-Y. Chen, H. Kinoshita, B. Zienicke, I. Molina and T. Lamparter, *Biochemistry*, 2009, **48**, 2817–27.
- 22 J. Dasgupta, R. R. Frontiera, K. C. Taylor, J. C. Lagarias and R. a. Mathies, *PNAS*, 2009, **106**, 1784–9.
- 23 M. E. Auldridge and K. T. Forest, *Crit. Rev. Biochem. Mol. Biol.*, 2011, **46**, 67–88.
- 24 P. Huo and D. F. Coker, *J. Phys. Chem. Lett.*, 2011, **2**, 825–833.
- 25 E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, **463**, 644–7.
- 26 G. Panitchayangkoon, D. Hayes, K. a. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship and G. S. Engel, *PNAS*, 2010, **107**, 12766–70.
- 27 D. S. Berns, H. L. Crespi and J. J. Katz, *J. Am. Chem. Soc.*, 1963, **85**, 8–14.
- 28 C. Romay, R. González, N. Ledón, D. Remirez and V. Rimbau, *Curr. Protein Pept. Sci.*, 2003, **4**, 207–16.
- 29 Y. Ren, J. Wan, X. Xu, Q. Zhang and G. Yang, *J. Phys. Chem. B*, 2006, **110**, 18665–9.
- 30 B. J. Homoelle and W. F. Beck, *Biochemistry*, 1997, **36**, 12970–5.
- 31 N. Adir, *Photosynth. Res.*, 2005, **85**, 15–32.
- 32 A. a. Arteni, L.-N. Liu, T. J. Aartsma, Y.-Z. Zhang, B.-C. Zhou and E. J. Boekema, *Photosynth. Res.*, 2008, **95**, 169–74.
- 33 L. David, A. Marx and N. Adir, *J. Mol. Biol.*, 2011, **405**, 201–13.
- 34 S. J. Grabowski, *Chem. Rev.*, 2011, **111**, 2597–625.
- 35 M. Röben, J. Hahn, E. Klein, T. Lamparter, G. Psakis, J. Hughes and P. Schmieder, *ChemPhysChem*, 2010, **11**, 1248–57.
- 36 J. Hahn, R. Kühne and P. Schmieder, *ChemBioChem*, 2007, **8**, 2249–55.
- 37 M. A. Mroginski, F. Mark, W. Thiel and P. Hildebrandt, *Biophys. J.*, 2007, **93**, 1885–94.
- 38 M. Röben and P. Schmieder, *Magn. Reson. Chem.*, 2011, **49**, 543–548.

-
- 39 T. Schirmer, W. Bode and R. Huber, *Journal of molecular biology*, 1987, **196**, 677–95.
- 40 M. Duerring, R. Huber and W. Bode, *FEBS Letters*, 1988, **236**, 167–170.
- 41 A. D. MacKerell, D. Bashford, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
- 42 A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–15.
- 43 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–802.
- 44 S. Kaminski and M. A. Mroginiski, *J. Phys. Chem. B*, 2010, **114**, 16677–86.
- 45 J. VandeVondele, M. Krack, F. Mohamedb, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.
- 46 T. Laino, F. Mohamed, A. Laio and M. Parrinello, *J. Chem. Theory Comput.*, 2005, **1**, 1176–1184.
- 47 T. Laino, F. Mohamed, A. Laio and M. Parrinello, *J. Chem. Theory Comput.*, 2006, **2**, 1370–1378.
- 48 C. Schiffmann and D. Sebastiani, *J. Chem. Theory Comput.*, 2011, **7**, 1307–1315.
- 49 A. C. Ihrig, C. Schiffmann and D. Sebastiani, *J. Chem. Phys.*, 2011, **135**, 214107.
- 50 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–99.
- 51 V. Weber, M. Iannuzzi, S. Giani, J. Hutter, R. Declerck and M. Waroquier, *J. Chem. Phys.*, 2009, **131**, 014106.
- 52 F. Jensen, *J. Chem. Theory Comput.*, 2008, **4**, 719–727.
- 53 J. Wan, X. Xu, Y. Ren and G. Yang, *J. Phys. Chem. B*, 2005, **109**, 11088–90.
- 54 U. Ryde and K. Nilsson, *J. Am. Chem. Soc.*, 2003, **125**, 14232–3.