IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2010 EPL 91 37004

(http://iopscience.iop.org/0295-5075/91/3/37004)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 152.78.199.38
The article was downloaded on 26/08/2010 at 15:20

Please note that terms and conditions apply.

A Letters Journal Exploring the Frontiers of Physics

August 2010

EPL, **91** (2010) 37004
doi: `10.1209/0295-5075/91/37004`

`www.epljournal.org`

Protein-protein interactions from linear-scaling first-principles quantum-mechanical calculations

D. J. Cole[1(a)], C.-K. Skylaris[2], E. Rajendra[3], A. R. Venkitaraman[3] and M. C. Payne[1]

[1] *Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge - J. J. Thomson Avenue, Cambridge CB3 0HE, UK, EU*
[2] *School of Chemistry, University of Southampton - Highfield, Southampton SO17 1BJ, UK, EU*
[3] *MRC Cancer Cell Unit Hutchison/MRC Research Centre - Hills Road, Cambridge CB2 0XZ, UK, EU*

received 30 March 2010; accepted in final form 23 July 2010
published online 23 August 2010

PACS `71.15.Mb` – Density functional theory, local density approximation, gradient and other corrections
PACS `87.15.km` – Protein-protein interactions
PACS `87.15.A-` – Theory, modeling, and computer simulation

Abstract – A modification of the MM-PBSA technique for calculating binding affinities of biomolecular complexes is presented. Classical molecular dynamics is used to explore the motion of the extended interface between two peptides derived from the BRC4 repeat of BRCA2 and the eukaryotic recombinase RAD51. The resulting trajectory is sampled using the linear-scaling density functional theory code, ONETEP, to determine from first principles, and with high computational efficiency, the relative free energies of binding of the ∼2800 atom receptor-ligand complexes. This new method provides the basis for computational interrogation of protein-protein and protein-ligand interactions within fields ranging from chemical biological studies to small-molecule binding behaviour, with both unprecedented chemical accuracy and affordable computational expense.

Copyright © EPLA, 2010

Protein-protein interactions (PPIs) are fundamental to all biological processes. The interrogation of PPIs amenable to chemical biological studies or small-molecule therapeutic intervention is an increasingly important field and, as such, new methods that provide insight into the behaviour of PPIs are of great interest. Computational methods that are capable of elucidating features of molecular recognition, such as PPI binding hotspots and binding affinities, are likely to drive experimental approaches to studying macromolecular complexes by defining the binding behaviour of peptides derived from existing structures and, more excitingly, structures that are not amenable to experimental structure determination. Existing approaches to calculating free energies of binding of protein complexes from force field based molecular dynamics (MD) vary in their degree of accuracy and computational expense. These range from methods such as free energy perturbation and thermodynamic integration [1], which follow a rigorous (but computationally very expensive) procedure, to approaches which rely on various approximations, such as the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method [2]. However, interactions at the typical length scales of protein-protein interfaces are governed by the laws of quantum mechanics (QM), so it would be more appropriate to perform calculations with first-principles QM approaches. Density Functional Theory (DFT) is the most widely used first-principles approach as it combines computational efficiency with the ability to calculate very accurate binding energies in, *e.g.*, biomolecules, as it describes, by construction, the electronic density distribution and its changes, such as charge transfer and polarisation, which take place upon association [3]. These crucial effects are not included in commonly employed classical force fields and, even though recent developments with polarisable force field models are promising, they are not as transferable or as accurate as first-principles approaches [4].

However, the applicability of DFT (and other first-principles approaches) to biomolecular assemblies is hindered by its computational expense, which in general is at least three orders of magnitude larger than force field methods, but most importantly by the scaling of the computational effort, which increases as the third (or greater) power of the number of atoms in the simulation. Several QM/MM approaches [5], which combine the

---

[(a)] E-mail: `djc56@cam.ac.uk`

accuracy of quantum mechanics in a small region with the efficiency of MM in a larger region, have been developed to address this computational bottleneck. These provide improved properties compared with pure MM approaches but they require great care as they suffer from the arbitrariness and effect of the coupling scheme between the two regions. Furthermore, they are not applicable in cases such as PPIs as the contact surface area can be as large as the proteins themselves. In this letter, we address this problem by using a new generation of DFT approach whose computational cost scales linearly with the number of atoms, allowing DFT calculations to be performed on entire protein-protein interfaces. In particular, we use the ONETEP [6] linear-scaling DFT program, which has been developed as a code for parallel computers [7] with novel and highly efficient algorithms that allow calculations for systems of tens of thousands of atoms [8]. ONETEP is unique as it achieves linear-scaling computational cost with plane-wave accuracy, meaning that the energy and other computed properties can be improved systematically by increasing a single parameter, equivalent to the energy cut-off in conventional plane-wave pseudopotential DFT approaches. Another point that we need to consider is that the interactions in biological systems are inherently dynamical and, hence, we need to combine large scale DFT calculations with methods that sample the conformational space spanned by our proteins. Towards this goal, we present a modification of the MM-PBSA technique for calculating the relative binding affinities of biomolecular complexes, which we call QM-PBSA.

The traditional MM-PBSA approach [2] computes free energy differences between bound and unbound states and has been used to study the relative binding affinities of many different systems including small molecules to proteins [9] and PPIs [10]. In the single trajectory approach [11], the MM-PBSA method requires a MD simulation of a receptor-ligand complex in a periodic box with explicit water, from which a set of representative structures are extracted. Water molecules and, usually, counter-ions are removed and the free energy of binding of the ligand to the receptor calculated according to

$$\Delta G_{MM} = \langle \Delta E_{MM} \rangle + \langle \Delta G_{PBSA}^{MM} \rangle - T\langle \Delta S_{MM} \rangle, \qquad (1)$$

where $\langle \Delta E_{MM} \rangle$ is the binding energy, usually a sum of van der Waals (vdW) and electrostatic (EL) terms from the MM force field, averaged over the ensemble of snapshots that are extracted from the MD simulation. Similarly, $\langle \Delta G_{PBSA}^{MM} \rangle$ is the binding free energy of solvation from the Poisson-Boltzmann continuum solvation model (including electrostatic solvation free energy (PB) and free energy due to surface area (SA)), averaged over the snapshots. Finally, $\langle \Delta S_{MM} \rangle$ is the gas phase binding entropy of the molecules which is usually computed from a normal mode analysis and averaged over the ensemble of snapshots. Often we are interested in relative free energies of binding of various ligands to the same receptor:

$$\begin{aligned}\Delta\Delta G_{MM} &= \Delta\langle \Delta E_{EL} \rangle + \Delta\langle \Delta E_{vdW} \rangle + \Delta\langle \Delta G_{PB} \rangle \\ &\quad + \Delta\langle \Delta G_{SA} \rangle - T\Delta\langle \Delta S_{MM} \rangle, \end{aligned} \qquad (2)$$

where the gas phase and solvation energies have been separated into polar and non-polar contributions.

In this work we introduce a new approach for PBSA calculations where, instead of using a classical force field to obtain the energy E_{MM} of each snapshot, we use a full DFT quantum-mechanical calculation. We will represent this energy as $E_{QM} = E_{DFT} + E_{disp}$, where E_{disp} is a dispersion correction [12] to the total DFT energy, E_{DFT}. For each snapshot we then include the solvation free energy as the weighted polar part of the solvation free energy from the MM calculation, plus the unweighted non-polar part, according to the formula

$$\Delta G_{PBSA}^{QM} = \Delta G_{PB} \times \left(\frac{\Delta E_{DFT}}{\Delta E_{EL}} \right)^{n_{PB}} + \Delta G_{SA}. \qquad (3)$$

n_{PB} is determined by a best-fit power law curve to a plot of ΔG_{PB} against ΔE_{EL} and is motivated by observations that complexes that are more strongly bound in the gas phase also tend to have higher solution phase binding free energies, which corresponds to $n_{PB} < 1$. We have made the assumptions that the polar part of the solvation free energy in QM-PBSA, ΔG_{PB}^{QM}, follows the same power law behaviour as its classical counterpart, ΔG_{PB}, and that, scaling of the short-range exchange-repulsion contributions, which are included in ΔE_{DFT} but not in ΔE_{EL}, by the exponent n_{PB} does not contribute significantly to the total QM free energy of binding. With this approach we are able to calculate differences in binding free energies in a fashion analogous to MM-PBSA, but where the gas phase energies now come directly from first-principles quantum mechanics:

$$\begin{aligned}\Delta\Delta G_{QM} &= \Delta\langle \Delta E_{DFT} \rangle + \Delta\langle \Delta E_{disp} \rangle \\ &\quad + \Delta\langle \Delta G_{PBSA}^{QM} \rangle - T\Delta\langle \Delta S_{MM} \rangle. \end{aligned} \qquad (4)$$

As a PPI test system, we consider here the interaction between the BRC4 repeat of BRCA2 and the protein RAD51, which is implicated, both *in vitro* and *in vivo*, in the error-free repair of double-stranded DNA breaks and hence necessary for the maintenance of genome stability and ultimately the prevention of cancer (reviewed in [13,14] and references therein). Figure 1 shows the high-resolution crystal structure of BRC4 (residues K1517-Q1551) bound to the core catalytic domain (E98-D339) of RAD51 [15]. BRC4 binds to RAD51 using a series of polar and non-polar contacts along the length of the peptide. It has recently been reported that all BRC repeats contain two binding hotspots, with distinct binding pockets in RAD51 [16], one containing the tetramer 1524-FHTA-1527 within a β-hairpin (BRC4A: P1519-K1533; no density is visible for the first two residues), and the second containing the tetramer 1545-LFDE-1548 within the context of an α-helix.

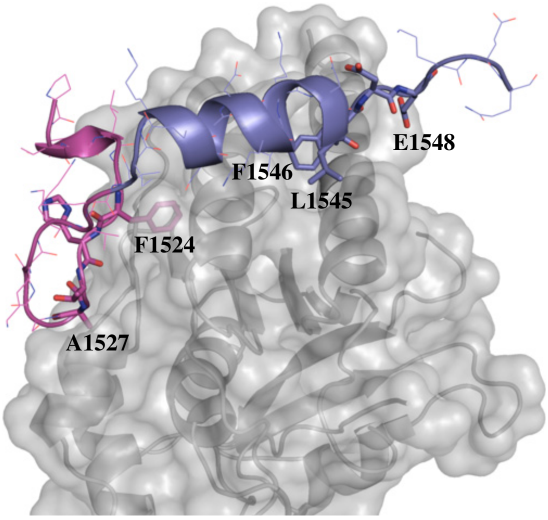

Fig. 1: The RAD51-BRC4 interface. The RAD51 receptor is shown in silver, the BRC4A peptide (P1519-K1533) in purple and BRC4B (I1534-Q1551) in blue. The FHTA and LFDE motif sidechains are depicted as sticks.

The dynamic behaviour of the RAD51-BRC4 complex was investigated using MD simulations, performed with the AMBER10 package [17]. Water molecules were treated using the TIP3P force field and all protein interactions were described by the AMBER ff99SB biomolecular force field [18]. Coulomb interactions were treated using the Particle Mesh Ewald sum, with a real space cut-off of 10 Å. The cut-off length for Lennard-Jones interactions was also set to 10 Å. Hydrogens were added to the X-ray crystal structure of the RAD51-BRC4 complex [15] (PDB: 1N0W), and a short energy minimisation was performed in vacuum to remove steric contacts. Water and sodium counter-ions were added and the system was heated to 300 K with weak harmonic restraints on the complex at constant pressure (NPT ensemble). Finally, all restraints were removed and the system was equilibrated for 2 ns at 300 K, at the end of which the root-mean-square deviation of the protein backbone atoms was converged and less than 2 Å relative to the original crystal structure. With a view to separating the energetic contributions to binding of the "FHTA" and "LFDE" motifs, we have removed all water molecules from the final equilibrated structure of RAD51-BRC4 and truncated the BRC4 peptide to include only the N-terminal 15 residues (P1519-K1533) that bind to the RAD51 oligomerisation interface (RAD51-BRC4A). Similarly, a RAD51-BRC4B system was created in which only the C-terminal residues of BRC4 (I1534-Q1551) were retained. RAD51 was terminated by $-\mathrm{NH}_3^+$ and $-\mathrm{COO}^-$ groups and the BRC4 half-peptides by $-\mathrm{NH}_3^+$ and $-\mathrm{CONH}_2$ in accordance with experimental procedure [16]. The two systems, RAD51-BRC4A and RAD51-BRC4B, were resolvated and, as above, were heated to 300 K and equilibrated for 2 ns. Production runs were performed, lasting 24 ns for each complex and snapshots were saved every 6 ps for MM-PBSA single trajectory analysis.

Table 1: Convergence of ΔG and $\Delta\Delta G_{QM}$ (kcal/mol) with the number of residues (N) in the RAD51 receptor included in the free energy calculation compared to the full system (210 residues) for one snapshot of each of the RAD51-BRC4A and RAD51-BRC4B systems. n_{PB} is calculated separately for each complex.

	RAD51-BRC4A		RAD51-BRC4B		
N	ΔG_{MM}	ΔG_{QM}	ΔG_{MM}	ΔG_{QM}	$\Delta\Delta G_{QM}$
33	−55.08	−63.05	–	–	–
69	−57.73	−66.97	–	–	–
163	−59.40	−65.57	−56.50	−68.46	−2.89
210	−59.48	−63.22	−56.87	−66.96	−3.74

Figure 2(a) shows typical snapshots from MD trajectories of the two systems. The BRC4A β-hairpin is stabilised by a number of intra-peptide hydrogen bonds, as reported in ref. [19]. The dominant RAD51-BRC4A inter-protein hydrogen bonds are shown in fig. 2(a) (Y191-H1525, D187-S1528 and D187-K1530) and contribute to binding in addition to hydrophobic interactions (most notably from F1524 and A1527 which reside in hydrophobic pockets in the receptor). The majority of the inter-protein hydrogen bonding in RAD51-BRC4B is found in the "LFDE" hotspot proposed in ref. [16] (R254-L1545 and a hydrogen-bonded network involving residues R247, R250, F1546 and E1548) and an additional hydrogen bond is formed between residues E213 and S1538. Although there are no large deviations from the crystal structure during the MD simulations, the fluctuations visible in fig. 2(b) justify our use of the dynamical trajectory as an input to the DFT calculation, rather than the static crystal structure. As is typical for biological systems, many of the inter-protein hydrogen bonds are intermittent over long time scales. For example, in the RAD51-BRC4A system, the Y191-H1525 backbone interaction is maintained throughout the simulation, while the D187-S1528 sidechain interaction is present for just 33% of the simulation and its contribution to the stability of the complex would therefore not be captured correctly by a static calculation.

We have performed both MM-PBSA and QM-PBSA analyses of the relative free energies of binding of the two half-peptides over the full 24 ns simulation, retaining 163 residues of the RAD51 receptor (a ∼2800 atom complex), which introduces errors of less than 1 kcal/mol (0.04 eV) in $\Delta\Delta G_{QM}$ (table 1). Classical free energy calculations were carried out using the MM-PBSA post-processing module in AMBER10. Infinite non-bonded cut-offs were used for the MM contributions. For calculating the free energy of solvation, dielectric constants of 1.0 and 80.0 were used for the solute and solvent, respectively, and the Poisson-Boltzmann equation was solved on a grid of spacing 0.5 Å. A spherical solvent probe of radius 1.4 Å and atomic radii provided by the AMBER force field were used for the implicit solvent molecules and solute atoms, respectively. The non-polar contribution to the free energy was calculated via $\Delta G_{SA} = \gamma SA$, where SA is the solvent-accessible

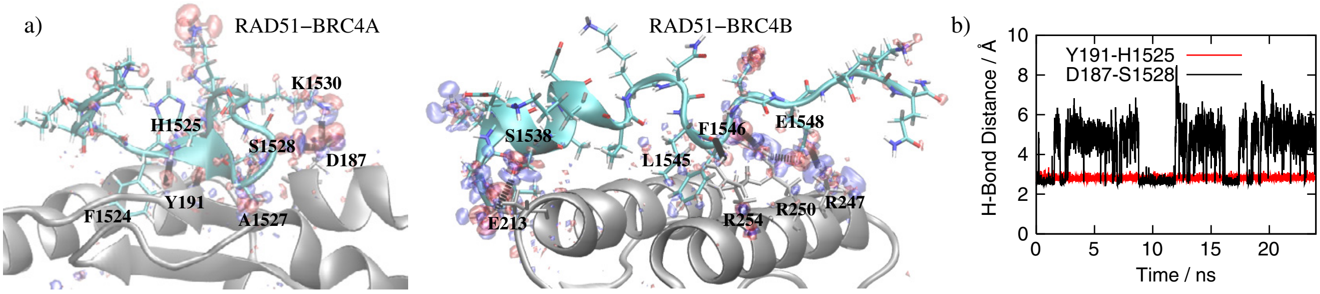

Fig. 2: (a) Snapshots from the RAD51-BRC4A and RAD51-BRC4B systems. The RAD51 receptor is shown in silver and inter-protein hydrogen bonds in black. Isosurfaces of electron density gain upon binding, obtained from QM calculations, are shown in red and loss in blue. (b) Inter-protein hydrogen bond distances as a function of time between selected residues.

surface area and γ is 0.0072 kcal/mol/Å^2. Entropic contributions to binding from the solute species, were estimated by normal mode analysis, using the NAB module of AMBER10. The trajectory was sampled every 0.75 ns and each snapshot was minimised in the generalised Born implicit solvent model, using first conjugate gradient and then Newton-Raphson minimisation, until the root-mean-square of the elements of the gradient vector was less than 10^{-10} kcal/mol/Å. The harmonic frequencies of the vibrational modes were calculated at 298 K for these minimised structures using normal mode analysis.

Quantum-mechanical calculations of total energies were performed with the ONETEP program [6], using the PBE gradient corrected exchange-correlation functional [20]. Interactions between electrons and nuclei were described by norm-conserving pseudopotentials. The ONETEP program achieves computational cost that scales linearly with the number of atoms by exploiting the "near-sightedness" of the single-particle density matrix $\rho(\mathbf{r},\mathbf{r}')$ in non-metallic systems. The density matrix is expressed in terms of a set of non-orthogonal generalised Wannier functions (NGWFs) [21] that are localised in real space:

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi^*_\beta(\mathbf{r}'), \tag{5}$$

and the total energy is self-consistently minimised with respect to both the density kernel $K^{\alpha\beta}$ and the NGWFs $\{\phi_\alpha(\mathbf{r})\}$. The NGWFs were expanded in a basis of periodic cardinal sine (psinc) functions [22]. To investigate the convergence of the gas phase binding energy ΔE_{QM} with psinc kinetic energy cut-off and NGWF radius, we have considered one snapshot of the interaction between the "FHTA" tetrapeptide and seven residues of RAD51 around the "FHTA" binding site (L186-A192). With NGWF radii of 4.0 Å, ΔE_{QM} is converged to 1.6 kcal/mol (0.07 eV), compared to radii of 4.8 Å. Using these NGWF radii, ΔE_{QM} is converged to within 0.3 kcal/mol (0.01 eV) with a psinc kinetic energy cut-off of 830 eV (when compared to a cut-off of 1020 eV). The spherical cut-off approach for Coulomb potentials [23] was used to eliminate all interactions of the molecules with their periodic images. Van der Waals interactions were included by augmenting the DFT energy expression by damped London potentials with parameters optimised specifically for the PBE functional [12]. No optimisation of the MM structures were performed prior to the energy calculations, which assumes a representative sampling of the potential energy landscape by the classical force field.

Table 2: Comparisons of mean absolute deviations (kcal/mol) of QM-PBSA (QM) and MM-PBSA (MM) from benchmark calculations for several protein complexes. Shown in brackets are the main interactions present, which are classed as charge-charge (c), polar (p) or hydrophobic (h). N is the number of snapshots sampled for each complex and Q_{rec} and Q_{lig} are the charges of the receptor and ligand, respectively (e$^-$).

Complex	N	Q_{rec}	Q_{lig}	QM	MM
LDNV-ASGK (cph)	2	−1	+1	1.5	2.5
LDN-SGK (cp)	5	−1	+1	4.7	3.0
LDN-AGK (c)	5	−1	+1	1.2	3.6
LDNV-AS (p)	4	−1	0	1.1	0.6
AYA-FH (h)	5	0	0	1.9	0.8

Before moving to the full RAD51-BRC4 systems, we first validate our solvation model using small fragments from the classical MD trajectory. Although it is straightforward to compare the accuracy of QM and MM gas phase binding energies, solution phase binding energy comparisons are made more difficult by the strong dependence of ΔG on the cavitation models used not only in PB calculations, but also in methods common in quantum chemistry, such as the polarisable continuum model (PCM). We therefore calculate the free energy of binding of several representative protein fragments using three methods, MM-PBSA, QM-PBSA and a direct DFT approach, each using identical PCM solvation models. Namely, a number of snapshots of small complexes (listed in table 2), containing a mixture of charged, polar and hydrophobic interactions, were extracted from the RAD51-BRC4A trajectory and terminated by –NH$_3^+$ and –COO$^-$ groups (fig. 3(a)). As a benchmark calculation, we compute directly the solution phase binding energy of each complex in Gaussian 09 [24] using the PBE exchange-correlation functional, a 6-311G* basis set, the default PCM [25] and the counterpoise correction for the

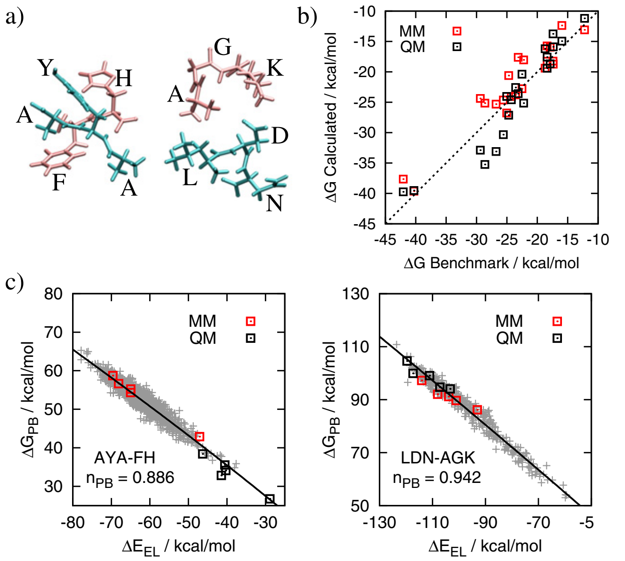

Fig. 3: (a) Examples of the AYA-FH and LDN-AGK complexes. (b) Correlation of ΔG calculated by MM-PBSA (red) and QM-PBSA (black) compared with benchmark results. (c) Determination of n_{PB} for two complexes from a power law fit to the classical ΔG_{PB} against ΔE_{EL} (grey crosses). For a small number of snapshots, ΔG_{PB} and ΔE_{EL} are explicitly calculated quantum mechanically (black) and compared to the original MM results (red).

basis set superposition error, using the method described in ref. [26]. All of these reference DFT energies were augmented by damped London potentials calculated using ONETEP [12]. The solution phase binding energy was also calculated using the charges and vdW parameters from the AMBER parm99SB force field and the default PCM in Gaussian 09. The correlation between the benchmark and MM results are shown in fig. 3(b) and the mean absolute deviation of the MM results from the benchmarks for each type of complex is listed in table 2. Interestingly, the MM force field appears to perform well for the smaller fragments for which it has been parameterised, but the deviations from the benchmark are in excess of 2 kcal/mol (0.09 eV) for the larger complexes.

As described earlier, n_{PB} was estimated from a best-fit line to a plot of ΔG_{PB} against ΔE_{EL} (fig. 3(c)) for the truncated complexes sampled 1000 times over the full 24 ns trajectories. By also plotting the benchmark gas phase binding energy and binding free energy of solvation on fig. 3(c), we can see that to a good approximation the QM data lies on the power law curves that we have fit to the MM data, despite the lack of polarisation in the latter. We therefore use eq. (3) and the DFT energy calculated within Gaussian 09 to scale the MM solvation energy and provide a QM-PBSA estimate to the total solution phase ΔG. Figure 3(b) reveals that overall there is good correlation between the QM-PBSA and benchmark calculations. The deviations between -25 and -30 kcal/mol are due to an overestimation of binding in the LDN-SGK complex, which is dominated by Coulomb interactions (table 2).

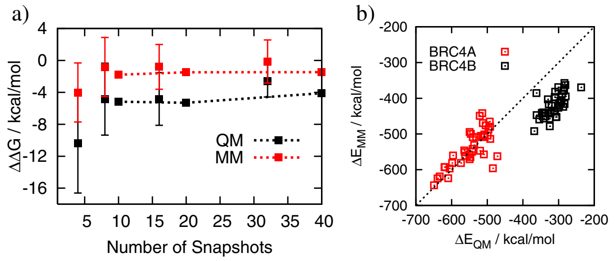

Fig. 4: (a) Convergence of $\Delta\Delta G = \Delta G(\text{RAD51-BRC4B}) - \Delta G(\text{RAD51-BRC4A})$ with the number of snapshots used in the QM/MM-PBSA analysis. Convergence is improved by correcting the evenly spaced sample set (points with error bars) to better sample the full distribution of energetic and dynamical properties (dashed lines). (b) Correlation between the QM and MM total gas phase binding energies. MM-PBSA overestimates the gas phase RAD51-BRC4B binding energy.

Interestingly, QM-PBSA performs better than MM-PBSA for the largest complex studied here (LDNV-ASGK), which contains a mixture of hydrophobic, polar and charged interactions, similar to typical protein-protein interfaces.

Turning now to the significantly larger RAD51-BRC4 complexes, and using ONETEP to calculate the DFT gas phase binding energies[1], we find that, in agreement with *in vitro* enzyme-linked immunosorbent assays [16], both BRC4A and BRC4B peptides display a similar ability to compete the RAD51-BRC4 interaction. Using a small set of equally spaced snapshots introduces a large random error in $\Delta\Delta G$ (fig. 4(a), points with error bars). However, given that there is a strong correlation between the QM and MM gas phase binding energies (fig. 4(b)), for every eight equally spaced snapshots, we also select two snapshots chosen to minimise the difference between properties of the sampled set (as calculated by MM) and the high sample limit of the MM distribution. The chosen properties are the mean and standard deviation of the free energy of binding in solution and the fractional occupancies of two intermittent hydrogen bonds (D187-S1528 and D187-K1530 for RAD51-BRC4A and R254-L1545 and R247-E1548 for RAD51-BRC4B). Using this sampling technique, the sum of the gas phase and solvation contributions to binding in QM-PBSA are converged to within 0.7 kcal/mol (0.03 eV) with respect to the number of snapshots sampled in both the BRC4A and BRC4B systems. Despite the fact that BRC4A and BRC4B bind to different sites on the RAD51 surface, the gas phase entropic binding energies of the two systems, estimated using normal modes analysis, are numerically identical in the two cases ($T\Delta S_{MM} = -39 \pm 1$ kcal/mol $(-1.69 \pm 0.04\,\text{eV})$). Hence, overall, QM-PBSA predicts that BRC4B is more strongly bound in solution by 4.1 kcal/mol (0.18 eV) (compared to 1.7 kcal/mol (0.07 eV) in MM-PBSA). Although there is very good agreement between QM and MM gas phase binding energies for BRC4A, this quantity is strongly overestimated for BRC4B in the MM approach (fig. 4(b)). This behaviour is similar to two of the complexes studied in the earlier solvation model validation. The gas phase binding energy of the uncharged AYA-FH complex was overestimated by the MM approach, leading to a small overestimation of the total solution phase binding energy, while the charged LDN-AGK complex was well represented by the MM force field in the gas phase. We speculate that for the BRC4A ligand, interactions are dominated by the Coulomb interactions between oppositely charged receptor and ligand, which are relatively easy to describe by point charge models, but for the uncharged BRC4B ligand, charge transfer and polarisation effects, which are not accounted for in the MM force field may become important. Examples of charge transfer at the two interfaces are shown in fig. 2(a). Finally, the average (maximum) magnitude of the forces on all the atoms in the RAD51-BRC4B complex are 18 kcal/mol/Å (140 kcal/mol/Å) in AMBER and 26 kcal/mol/Å (200 kcal/mol/Å) in ONETEP, which implies that no unphysical conformations are being accessed by the classical force field in this system.

---

[1] A full QM simulation of a 2800 atom complex requires ∼1000 CPU hours on 2.4 GHz Intel Nehalem processors.

In conclusion, we have presented a new approach that allows the study of macromolecular complexes consisting of thousands of atoms within DFT, opening the way for a whole new level of understanding of biologically important systems, such as the behaviour of binding hotspots at protein interfaces. First-principles approaches are, by their nature, fully transferable to any biomolecular complex, including PPIs, small-molecule ligands and systems containing transition metal ions, because they include charge transfer and polarisation effects. Furthermore, the DFT approach provides all of the electronic structure information, for example optical absorption and electronic transport properties [27], that have been used extensively as experimental probes in areas such as materials science and biochemistry.

* * *

We would like to thank W. Belfield for useful discussions. Computational resources were provided by the Cambridge HPC Service, funded by EPSRC Grant EP/F032773/1, and by the IRIDIS3 supercomputer of the University of Southampton. DJC is supported by the EPSRC. CKS would like to thank the Royal Society for a University Research Fellowship. ER is supported by a PhD studentship from the UK Medical Research Council, which supports the work in ARV's laboratory.

REFERENCES

[1] Deng Y. and Roux B., *J. Phys. Chem. B*, **113** (2009) 2234.

[2] Srinivasan J., Cheatham III T. E., Cieplak P., Kollman P. A. and Case D. A., *J. Am. Chem. Soc.*, **120** (1998) 9401.

[3] Heady L., Fernandez-Serra M., Mancera R. L., Joyce S., Venkitaraman A. R., Artacho E., Skylaris C. K., Colombi Ciacchi L. and Payne M. C., *J. Med. Chem.*, **49** (2006) 5141.

[4] Ponder J. W. *et al.*, *J. Phys. Chem. B*, **114** (2010) 2549.

[5] Thiel W., in *Multiscale Simulation Methods in Molecular Sciences*, edited by Grotendorst J., Attig N., Blügel S. and Marx D., Vol. **42** (Jülich Supercomputing Centre, Jülich) 2009, pp. 203–214, `http://www.fz-juelich.de/nic-series/NIC-Series-e.html`.

[6] Skylaris C. K., Haynes P. D., Mostofi A. A. and Payne M. C., *J. Chem. Phys.*, **122** (2005) 084119.

[7] Skylaris C. K., Haynes P. D., Mostofi A. A. and Payne M. C., *Phys. Status Solidi B*, **243** (2006) 973.

[8] Hine N. D. M., Haynes P. D., Mostofi A. A., Skylaris C. K. and Payne M. C., *Comput. Phys. Commun.*, **180** (2009) 1041.

[9] Wang J., Morin P., Wang W. and Kollman P. A., *J. Am. Chem. Soc.*, **123** (2001) 5221.

[10] Gohlke H., Kiel C. and Case D. A., *J. Mol. Biol.*, **330** (2003) 891.

[11] Massova I. and Kollman P. A., *J. Am. Chem. Soc.*, **121** (1999) 8133.

[12] Hill Q. and Skylaris C. K., *Proc. R. Soc. London, Ser. A*, **465** (2009) 669.

[13] West S. C., *Nat. Rev. Mol. Cell Biol.*, **4** (2003) 435.

[14] Venkitaraman A. R., *Annu. Rev. Pathol.*, **4** (2009) 461.

[15] Pellegrini L., Yu D. S., Lo T., Anand S., Lee M., Blundell T. L. and Venkitaraman A. R., *Nature*, **420** (2002) 287.

[16] Rajendra E. and Venkitaraman A. R., *Nucl. Acids Res.*, **38** (2009) 82.

[17] Case D. A. *et al.*, Amber10, University of California, San Francisco (2008).

[18] Hornak V., Abel R., Okur A., Strockbine B., Roitberg A. and Simmerling C., *Proteins*, **65** (2006) 712.

[19] Buis N., Skylaris C. K., Grant G. H., Rajendra E., Payne M. C. and Venkitaraman A. R., *Mol. Simul.*, **34** (2008) 749.

[20] Perdew J. P., Burke K. and Ernzerhof M., *Phys. Rev. Lett.*, **77** (1996) 3865.

[21] Skylaris C. K., Mostofi A. A., Haynes P. D., Dieguez O. and Payne M. C., *Phys. Rev. B*, **66** (2002) 035119.

[22] Mostofi A. A., Haynes P. D., Skylaris C. K. and Payne M. C., *J. Chem. Phys.*, **119** (2003) 8842.

[23] Jarvis M. R., White I. D., Godby R. W. and Payne M. C., *Phys. Rev. B*, **56** (1997) 14972.

[24] Frisch M. J. *et al.*, Gaussian 09, Gaussian Inc., Wallingford, CT (2009).

[25] Tomasi J., Mennucci B. and Cammi R., *Chem. Rev.*, **105** (2005) 2999.

[26] Riley K. E., Vondrásek J. and Hobza P., *Phys. Chem. Chem. Phys.*, **9** (2007) 5555.

[27] Artacho E., Machado M., Sánchez-Portal D., Ordejón P. and Soler J. M., *Mol. Phys.*, **101** (2003) 1587.