J Comput Aided Mol Des (2011) 25:1071–1084
DOI 10.1007/s10822-011-9484-3

Computational and experimental studies of the interaction between phospho-peptides and the C-terminal domain of BRCA1

Victor M. Anisimov · Arturas Ziemys · Smitha Kizhake · Ziyan Yuan · Amarnath Natarajan · Claudio N. Cavasotto

Received: 3 August 2011 / Accepted: 2 November 2011 / Published online: 16 November 2011

Abstract The C-terminal domain of BRCA1(BRCT) is involved in the DNA repair pathway by recognizing the pSXXF motif in interacting proteins. It has been reported that short peptides containing this motif bind to BRCA1(BRCT) in the micromolar range with high specificity. In this work, the binding of pSXXF peptides has been studied computationally and experimentally in order to characterize their interaction with BRCA1(BRCT). Elucidation of the contacts that drive the protein–ligand interaction is critical for the development of high affinity small-molecule BRCA1 inhibitors. Molecular dynamics (MD) simulations revealed the key role of threonine at the peptide P+2 position in providing structural rigidity to the ligand in the bound state. The mutation at P+1 had minor effects. Peptide extension at the N-terminal position with the naphthyl amino acid exhibited a modest increase in binding affinity, what could be explained by the dispersion interaction of the naphthyl side-chain with a hydrophobic patch. Three in silico end-point methods were considered for the calculation of binding free energy. The Molecular Mechanics Poisson–Boltzmann Surface Area and the Solvated Interaction Energy methods gave reasonable agreement with experimental data, exhibiting a Pearlman predictive index of 0.71 and 0.78, respectively. The MM-quantum mechanics-surface area method yielded improved results, which was characterized by a Pearlman index of 0.78. The correlation coefficients were 0.59, 0.61 and 0.69, respectively. The ability to apply a QM level of theory within an end-point binding free energy protocol may provide a way for a consistent improvement of accuracy in computer-aided drug design.

Keywords Quantum mechanics · Semiempirical Hamiltonian · MM-QMSA (MM/QM-COSMO) · BRCA1 · End-point binding free energy methods

Claudio N. Cavasotto—Address after January 2012: Instituto de Biomedicina de Buenos Aires - Max Planck Society Partner (IBioBA - MPSP), Godoy Cruz 2201, C1425FQA, Buenos Aires, Argentina.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9484-3) contains supplementary material, which is available to authorized users.

V. M. Anisimov · A. Ziemys · C. N. Cavasotto (✉)
School of Biomedical Informatics, University of Texas Health Science Center at Houston, 7000 Fannin Ste. 690, Houston, TX 77030, USA
e-mail: cnc@cavasotto-lab.net;
Claudio.N.Cavasotto@uth.tmc.edu

Present Address:
A. Ziemys
The Methodist Hospital Research Institute, 6670 Bertner Ave., Houston, TX 77030, USA

S. Kizhake · Z. Yuan · A. Natarajan
Chemical Biology Program, Department of Pharmacology and Toxicology, University of Texas Medical Branch, Galveston, TX 77555, USA

Present Address:
Z. Yuan
Kansas University Medical Center, Mailstop 4049, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA

S. Kizhake · A. Natarajan (✉)
Eppley Institute for Cancer Research, University of Nebraska Medical Center, Omaha, NE 68198, USA
e-mail: anatarajan@unmc.edu

Introduction

Mutations in the breast cancer susceptibility gene 1 (BRCA1) predispose women to breast and ovarian cancers. The BRCA1

gene product is a 220-kDa protein that has functional domains at its N and C termini [1]. Protein–protein interactions mediated by the C-terminus domains of BRCA1(BRCT) regulate a number of DNA damage response functions. In response to DNA damage, proteins CCDC98/Abraxas, BACH1 and CtIP recognize and bind the BRCT(BRCA1) to regulate cell cycle checkpoints and activate DNA repair [2–7]. Preclinical studies suggest that cells with mutations in BRCT(BRCA1) are sensitive to DNA damage based therapeutics [8]. Therefore, use of small molecule inhibitors that functionally mimic these chemosensitizing mutations may be an attractive therapeutic strategy. Towards this goal we have identified short phosphoserine peptides (pSXXF) that bind BRCT(BRCA1) with low-micromolar affinities [9]. Structural and biochemical studies have shown that the peptides bind BRCT(BRCA1) in a two-point binding mode. The phosphate group on the serine residues makes hydrophilic contact with the backbone and side chain residues on the BRCT(BRCA1). The phenyl ring at the P+3 position resides in a hydrophobic pocket $\sim$15 Å away from the hydrophilic contact [10–17].

Structural studies show that (1) the oxygen atoms on the phosphate group makes three hydrogen bonds with the backbone nitrogen of G1655, the side chain oxygen of S1656 and the ε-nitrogen of K1702 and (2) the P+3 phenyl alanine sits in a hydrophobic pocket made by F1704, M1775 and N1774 [10–17] (cf. Fig. 1). To further dissect this interface we recently generated a library of tetrapeptides and evaluated their binding ability to BRCT(BRCA1) using a set of biochemical techniques. The results from this study suggests that all four residues contribute to BRCT(BRCA1) with the phosphoserine and phenylalanine (P+3) residue being the major contributors. More importantly even small changes to the tetrapeptide results in a loss of activity which is an indication of binding specificity [18]. A recent report also showed that BRCT(BRCA1) is the more selective for a phosphoserine as opposed to a phosphotheronine by more than 2 orders of magnitude, while the BRCT domains of TopBP1 shows no such selectivity [19].

Analysis of the BRCT(BRCA1)-phosphopeptide structures show several hydrophobic patches/crevices around

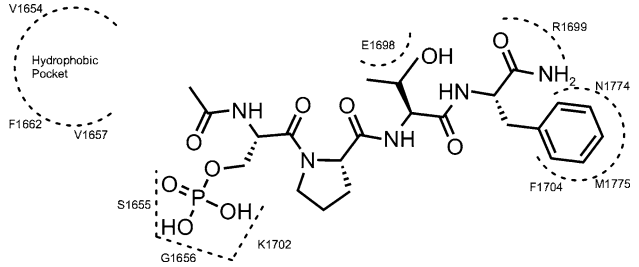

Fig. 1 Interactions of T1 (Ac-pSPTF-$CONH_2$) with BRCT(BRCA1) (from PDB 1T29)

the pSXXF binding site. These additional binding sites can be effectively exploited to improve the binding affinity for the target protein. One such binding site on BRCT(BRCA1) is found at the N-terminus of the pSXXF binding site formed by P1659, V1654, L1657 and F1662 (PVLF) residues. We hypothesize that conjugating a hydrophobic moiety (that occupies the PVLF cluster) to the N-terminus of the parent peptide pSPTF will lead to increased binding affinity for BRCT(BRCA1).

In the first part of this study, we synthesized pentapeptide P7 (Ac-Nap-pSPTF-$CONH_2$), which includes a hydrophobic naphthyl functional group that could potentially occupy the PVLF cluster. In the second part of this study we performed MD simulations to further characterize the interaction of BRCA1(BRCT) with a family of pSXXF phosphopeptides and P7, in order to assess the impact of point mutations in binding affinity and the dynamics of the system; the importance of properly accounting for flexibility in studying protein–ligand interaction is already solidly documented [20, 21]. In the third part of the study, the binding free energy between the phosphopeptides and BRCA1(BRCT) was calculated in silico with three end-point methods. We were seeking to obtain an adequate balance of accuracy and computational speed. End-point methods are much more computationally affordable than thermodynamic integration [22] (TI) and free energy perturbation [23] (FEP) method. We considered two classical-mechanics based methods, the molecular mechanics/Poisson Boltzmann-surface area [24, 25] (MM/PB-SA), and the MM/PB-SA-derived solvated interaction energy (SIE) [26] method. Completing the list of simulation methods, we used the recently developed Molecular Mechanics/Quantum Mechanics-COnductor-like Screening Model [27] (MM/QM-COSMO) in which rescoring of the classical MD trajectories is performed using a QM representation of the complete system and a continuum solvent model; an enhanced method to calculate entropic change based on configurational integrals is also used. For the sake of simplicity, we will refer from now on to this method as MM-quantum mechanics-surface area (MM-QMSA). The MM-QMSA method was recently successfully applied to calculate binding free energies of phosphopeptides to the SH2 domain of human Lck p56 tyrosine kinase, adding to the list of recent applications of QM methods to study ligand binding [28–32].

Semiempirical methods have the advantage that they explicitly describe the effects of charge transfer and polarization, which are essential part of protein–ligand interaction [33, 34]. Electronic polarization is responsible for the deformation of charge distribution due to an external electric field. Being formulated on the quantum mechanical framework, semiempirical methods provide greater degree of parameter transferability than it is

possible with classical force fields. Due to the significant progress in linear scaling quantum chemistry and computer hardware, semiempirical QM calculations of biological systems are routinely affordable now. In this contribution, we continue the series of publications in which we utilize the advantage of using semiempirical methods to improve the accuracy of protein–ligand binding free energy prediction [27].

Materials and methods

The phosphorylated peptides used in the current study were Ac-pSPTF-$CONH_2$ (T1), Ac-pSATF-$CONH_2$ (T2), Ac-pSPAF-$CONH_2$ (T3), Ac-EPTF-$CONH_2$ (T4), Ac-pSPVF-$CONH_2$ (T5), and Ac-Nap-pSPTF-$CONH_2$ (P7), where pS stands for phosphorylated amino acid serine, and Nap for the non-natural 2-naphthyl amino acid (Fig. 2). A list of these peptides is shown in Table 1.

Sample preparation

The plasmid construct (pAM15, gift from Luc Gadraeu, University De Sherbrook) encoding six his-tagged BRCA1(BRCT) domains of BRCA-1 (amino acids 1646-1859) was used to transform BL21(DE3) RIL (Stratagene). Protein expression was induced by 1 mM IPTG and the recombinant protein was purified by nickel affinity chromatography (Qiagen). The protein was concentrated tenfold using a molecular weight cut-off centricon filter (Millipore) at 4 °C. The concentrated protein sample was then filtered using a 2 micron filter (Millipore). Homogeneity of the purified protein preparation was assessed by SDS–PAGE. The filtered samples were dialyzed 3 times, for 2 h each, against phosphate buffer (50 mM KH_2PO_4 pH = 7.2) at room temperature. The concentration of the protein sample was determined by the BCA method using a Thermo-Pierce kit (Thermo Scientific Pierce BCA protein assay). All peptides used in this study were synthesized and purified by the Tufts core facility. The solid peptides were reconstituted in water and amino acid analysis was done to determine concentrations. Peptide solutions for the ITC study were made by freeze-drying the aqueous solutions and reconstituting the resulting solid in 50 mM KH_2PO_4 pH = 7.2.

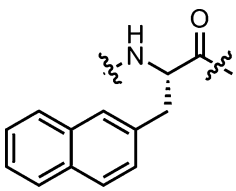

Fig. 2 Structure of the non-natural naphthyl amino acid (Nap)

Table 1 Composition and binding free energy of T1–T5 and P7 peptides

Peptide	ID	ΔG_{exp} (kcal/mol)	Refs.
Ac-pSPTF-$CONH_2$	T1	−7.5	[9]
Ac-pSATF-$CONH_2$	T2	−6.7	[9]
Ac-pSPAF-$CONH_2$	T3	−5.9	[9]
Ac-EPTF-$CONH_2$	T4	No binding	[9]
Ac-pSPVF-$CONH_2$	T5	−7.8	[18]
Ac-Nap-pSPTF-$CONH_2$	P7	−7.8	This work

Free energy measurements

A MicroCal VP-ITC instrument was used to study the thermodynamic factors arising from the protein-peptide interaction. All experiments were carried out at 25 °C. The reference cell was filled with buffer (50 mM KH_2PO_4 pH = 7.2) remaining after the third dialysis. The reference power supplied to the cells was 15 μcal/s. Varying concentrations of the peptide were loaded in the syringe (100–900 μM) and 10–30 μM of His-BRCT was loaded in the cell. All solutions were filtered through a 2 micron filter before loading into syringe or cell. Samples were also degassed with stirring for 10 min at 25 °C using the Thermovac.

The first injection of 3 μL of peptide into protein was followed by 29–10 μL injections spaced every 3 min apart. The solution was constantly stirred at 307 rpm. Heats in the range of 0.2–0.5 μcal/s were seen for most peptides. Peptide dilution curves were obtained by titrating the peptide into buffer under similar conditions. This curve was subtracted from the sample curve to get the corrected heats obtained for the reaction. The data obtained was integrated using Origin 7 software supplied with the MicroCal VP-ITC.

Molecular dynamics

Molecular dynamics simulations were performed on the C-terminal domain of the BRCA1 protein (PDB code 1T29). The protonation state of His, as well as flips of Gln and Asn amino acids, were visually inspected. Hydrogen atoms were added at physiological condition by using AMBER package *leap* module. After that, the protein acquired a net charge of −5 electron units. The crystallographic bound decapeptide was used as a template to prepare the starting geometries of the T1–T5 and P7 phosphorylated peptides. The peptides were terminated by acetyl and *N*-methyl groups, as standard in AMBER. This resulted in replacing the C-terminal amide by a *N*-methyl amide, what has a negligible effect on the outcome. The peptides carried a net charge of −2 electron units due to the phosphate group. The peptide T4 carried a charge of −1. Three distinct manually docked conformations of P7 were examined. In the first conformation the side chain

of Nap was positioned toward amino acids L1657, T1658, P1659. In the second conformation the Nap side chain was arranged in a small pocket formed by amino acids M1689, K1690, T1691, C1697, and E1698. The third examined conformation had the naphthyl group directed toward the solvent; however during the MD simulation it spontaneously converged to the first conformation. Out of the two stable conformations the first one showed more favorable binding free energy, therefore it was retained, and the second and third conformations of P7 were excluded from the analysis.

Molecular dynamics simulations were performed using AMBER ff99SB force field [35]. The missing parameters for Nap were derived using REDDB parameter development tool [36] (The AMBER parameter file for P7 is displayed in Supplementary Material). The AMBER parameters for pS were obtained from the work of Homeyer et al. [37]. Simulations were performed using the NAMD 2.6 [38] program, and an isothermal-isobaric NPT ensemble under periodic boundary condition. The protein–ligand complex was placed in cubic box of TIP3P water molecules [39] with a minimum of 13 Å solvent padding on each side and neutralized by adding sodium counterions by using the AMBER *leap* utility. A non-bonded cutoff of 12 Å was used, and the non-bonding interactions were switched on starting at 10 Å. The SHAKE algorithm [40] was applied to constrain bonds involving hydrogen atoms. The system geometry was minimized for 1,000 steps under a harmonic restraint of 10 kcal/(mol * Å^2) placed on the coordinates of the complex backbone atoms. Holding the restraints, the system was gradually heated to 300 K within 30 ps and then equilibrated for another 50 ps after removing the restrains. Constant temperature was maintained by using a stochastic Langevin thermostat with a collision frequency of 1 ps^{-1}. A constant pressure of 1 atm was maintained by using the Langevin piston method with a period of 200 fs and a damping timescale of 100 fs. Long-range electrostatics was treated by the particle-mesh Ewald method using grid density of about 1 Å^{-1}. The integration time step was set to 2 fs and the dynamics trajectory was saved with 1 ps interval. Each complex was simulated for 10 ns. The NAMD generated trajectories were converted to AMBER format by using *ptraj* utility [41]. The first 5 ns of the NPT simulation were treated as equilibration, and the remaining 5 ns were considered as production trajectory.

MM/PB-SA free energy calculations

Protein–ligand binding affinities ΔG_{bind} were computed using the MM/PB-SA method, where the binding free energy is calculated as the difference between the bound and unbound states of protein and ligand, according to [24, 25, 27, 42, 43]

$$\Delta G_{\text{bind}} = \Delta\langle E_{\text{MM}}\rangle + \Delta G_{\text{solv}} - T\Delta S_{\text{bind}} \quad (1)$$

where $\langle\cdots\rangle$ is the ensemble average, E_{MM} is the gas-phase potential energy, $\Delta G_{\text{solv}} = \Delta\langle W\rangle$ (W is the effective solvation energy which incorporates the solvent degrees of freedom [DOF]), and the third term accounts for the entropy change of the solute. The entropy change of the solvent is implicitly taken into account in ΔG_{solv}.

Binding free energy calculations were performed averaging over 1,000 snapshots extracted from the production trajectory with 5 ps interval, after water, ions, and periodicity were removed. This separation is enough to consider uncorrelated motions and thus to report errors as the variance of energy averages [44]. Calculations were performed using a single-trajectory approach where unbound protein and ligand conformations are extracted from the complex trajectory. The single-trajectory approach has been shown to reduce noise in end-point calculations [45]. The gas-phase molecular mechanics energy term (E_{MM}) was calculated using the AMBER *sander* module [41] treating all non-bonded interactions explicitly, and with an internal dielectric constant of 1. There were no conserved water molecules in the protein–ligand area observed during the MD simulation, and therefore all water molecules were deleted for the purpose of free energy calculation.

The solvation free energy ΔG_{solv} was separated into polar and non-polar contributions as

$$\Delta G_{\text{solv}} = \Delta G^{\text{pol}}_{\text{solv}} + \Delta G^{\text{np}}_{\text{solv}} \quad (2)$$

The polar contribution was calculated according to the PBSA method using the *pbsa* utility in AMBER [41]. The dielectric constant for water was taken as 80. The calculations were performed using atomic radii specifically optimized for Poisson–Boltzmann implicit solvent model by Swanson et al. [46]. The missing radii were derived by analogy with the published atom types given in the form of AMBER atom names, and were assigned the following values (in Å): CH3(Ace) = 2.595; CZ*(Trp) = 1.651; CH2(Trp) = 1.651; CD(Glu) = 1.942; OXT(C-terminal) = 1.516; H4, H5(His, Trp) = 0.840, CB(Nap) = 2.087, CG(Nap) = 1.651, C*(Nap) = 1.837, where the amino acid name is shown in brackets and "*" indicates wildcard in the atom name. The following radii for the phosphate group were taken from the work of Tsui et al. [47] (in Å): OG = 1.500, P = 1.850, O1P = 1.500, O2P = 1.500, and O3P = 1.500.

The non-polar contribution, which accounts for the energy to build the cavity and the dispersion interaction, was determined using the solvent-accessible-surface-area (SASA) approach,

$$G_{\text{nonp}} = \gamma\,\text{SASA} \quad (3)$$

where the default value of the surface tension coefficient γ was used [0.0072 kcal/(mol * Å^2)]. The surface area was computed with the *molsurf* program in AMBER, using a solvent probe radius of 1.4 Å.

The entropic contribution ΔS_{bind} to protein–ligand binding free energy was computed as a sum of rigid-body rotational ΔS_{rot} and translational ΔS_{tran}, and vibrational ΔS_{vib} terms according to

$$\Delta S_{\text{bind}} = \Delta S_{\text{rot}} + \Delta S_{\text{tran}} + \Delta S_{\text{vib}} \quad (4)$$

Assuming that the ligand rotational and translational DOF are uncorrelated and decoupled from the protein motion, the rotational and translational entropy terms are computed using the following equations [27, 48, 49]

$$\Delta S_{\text{rot}} = \frac{3}{2}R + R\ln\left[\frac{1}{8\pi^2}(2\pi)^{3/2}\sqrt{\langle\Delta\alpha^2\rangle\langle\Delta\beta^2\rangle\langle\Delta\gamma^2\rangle}\right] \quad (5)$$

$$\Delta S_{\text{tran}} = \frac{3}{2}R + R\ln\left[C^0(2\pi)^{3/2}\sqrt{\langle\Delta x^2\rangle\langle\Delta y^2\rangle\langle\Delta z^2\rangle}\right] \quad (6)$$

where R is universal gas constant, C^0 is the standard concentration of 1 M or 1 molecule per 1,660 Å^3, Δx^2, Δy^2, Δz^2 are the variances of the ligand center of mass from its average position when the complex snapshots are aligned by protein atoms, and $\Delta\alpha^2$, $\Delta\beta^2$, $\Delta\gamma^2$ are the variances of ligand rotation angles around the x, y, and z axes in the protein frame. The rigid-body entropy loss of the ligand was computed for 5 independent windows each containing 200 snapshots along the MD trajectory using the algorithm described by Swanson et al. [48], and block-averaging the computed values. The vibrational entropy change ΔS_{vib} upon complex formation was computed using AMBER's *nmode* utility as the difference in vibrational entropy of complex and its constituents within the harmonic oscillator approximation.

$$\Delta S_{\text{vib}} = S_{\text{vib}}^{\text{cpx}} - S_{\text{vib}}^{\text{pro}} - S_{\text{vib}}^{\text{lig}} \quad (7)$$

The snapshot geometry for the purpose of normal mode analysis was minimized to meet a RMS gradient of 0.0001 kcal/(mol Å) using a distance-dependent dielectric constant of $4r$. The vibrational entropy was averaged over 1,000 snapshots. The conformational entropy change corresponding to the loss of dihedral DOF upon complexation was not considered due to its lesser significance, especially for the calculation of relative binding free energy of series of congeneric ligands [27]. This was confirmed by a Monte Carlo simulation with minimization [50] of T1 using a continuum solvation model, similar to what has been presented recently [27] (see Refs. [51–53] for computational details). An upper bound for the conformational entropy change could be obtained assuming that the low energy conformations within 5 kcal/mol were equally visited, and that the bound state were characterized by a single conformation, thus obtaining $RT\ln N_{\text{conf}}$ ~2.3 kcal/mol. This value is smaller than the rigid body and vibrational entropies (see "Result" section). However, only the lowest energy states will be heavily populated –thus reducing the actual number of distinct accessible conformational states, and the bound ligand might keep some residual entropy. So, the actual change in conformational entropy would be smaller than 2.3 kcal/mol.

SIE scoring calculations

The binding affinity was also computed using the SIE empirical scoring function with its parameters being fitted to experimental binding affinity of 99 protein–ligand complexes [26],

$$\Delta G_{\text{bind}}(\alpha, \varepsilon_{\text{in}}, \rho, \gamma, C) = \alpha[E_{\text{vdw}} + E_{\text{coul}}(\varepsilon_{\text{in}}) + \Delta E_{\text{rf}}(\rho, \varepsilon_{\text{in}}) + \gamma\Delta\text{MSA}(\rho)] + C \quad (8)$$

The SIE function is internally based on the standard MM/PB-SA approach with additional empirical scaling applied to the energy terms. Calculations were performed using the *sietraj* utility in the SIE program package with default parameters [54] by postprocessing 1,000 snapshots of protein–ligand complexes written in the format of AMBER trajectory file after water, ions, and periodicity were removed. The reaction field contribution (ΔE_{rf}) was obtained by solving the Poisson–Boltzmann equation by an integrated solver employing the AMBER force field ff03 [55, 56] parameters with atomic radii scaled by factor $\rho = 1.1$, and using an internal dielectric constant $\varepsilon_{\text{in}} = 2.25$, and an external dielectric constant $\varepsilon_{\text{out}} = 78.5$. The non-polar contribution was computed on molecular surface area (MSA) by *sietraj* tool and using surface tension coefficient $\gamma = 0.012894$ kcal/(mol*Å^2). The computed sum of internal energy ($E_{\text{vdw}} + E_{\text{coul}}$), the reaction field ($\Delta E_{\text{rf}}$), and the non-polar contribution ($\gamma\Delta$MSA) were scaled by a factor $\alpha = 0.104758$. The result was augmented by a constant $C = -2.89$ kcal/mol, thus producing the final value of the binding free energy. No explicit entropy term is included in the SIE formalism which relies on enthalpy-entropy cancellation.

MM-QMSA calculations of binding free energy

Quantum mechanics calculations were performed according to the MM-QMSA approach [27], where the binding free energy is calculated in an analogue way to Eq. 1 as

$$\Delta G_{\text{bind}} = \langle\Delta H^{\text{COSMO}}\rangle - T\Delta S_{\text{bind}} \quad (9)$$

where H^{COSMO} includes the QM energy of solute in solvated state (calculated using the COSMO continuum solvent model [57], E^{COSMO}) and the non-polar contribution to the solvation energy according to

$$H^{\text{COSMO}} = E^{\text{COSMO}} + G_{\text{solv}}^{\text{np}} \quad (10)$$

The term $\Delta H^{\mathrm{COSMO}}$ represents the enthalpy difference between the complex and its constituents, and it also accounts for the entropic change of the solvent. The entropy term $T\Delta S_{\mathrm{bind}}$ in Eq. 9 was computed as described above for MM/PB-SA.

The QM-COSMO calculations were performed on the same structures which were used in MM/PB-SA and SIE calculations. Complex, protein, and ligand were subjected each to 100 steps of geometry relaxation using the continuum solvent model COSMO while treating the entire system at quantum mechanical level. Performing a longer relaxation may lead to structure deviation from the experimental geometry [58, 59], which is related to insufficient optimization of semiempirical Hamiltonians for condensed phase. This is not surprising considering the novelty of the application of semiempirical methods to macromolecular systems. Hence, in order to only relieve the structural strain we performed a limited geometry optimization as was previously done by us [27] and others [60].

Quantum mechanics calculations were performed by using the LocalSCF program package [61] with implicit solvent model COSMO and the PM3 Hamiltonian [62] (cf. Refs. [27, 63] for details). The COSMO cavity surface was constructed as a union of atomic van der Waals radii which were previously optimized to fit hydration free energy of 98 small molecules [27]. The surface charges were generated utilizing 32 triangular segments assigned to non-hydrogen atoms and 12 segments assigned to hydrogen atoms. The non-polar contribution to hydration free energy was computed for solvent accessible surface using a solvent probe radius of 1.3 Å, and a surface tension coefficient set to 0.0020 kcal/(mol * Å^2) as previously described [27]. The exchange interactions and the product of Fock matrix with localized molecular orbital vectors were computed using a 6 Å cutoff. Empirical corrections were applied to maintain the planarity of peptide group (keyword MMOK) and to correct the geometry of 5- and 6-membered saturated rings (keyword MMCCROK). Long-range electrostatics was treated via the fast multipole method under default program settings.

In semiempirical QM theory electron-repulsion integrals implicitly include the effect of dispersion interactions via model parameters [64]. Nevertheless, the traditional semiempirical Hamiltonians underestimate the strength of mid-range dispersion interactions [65]. Improving this term would require a complete a reparameterization of the Hamiltonians. A simpler but still efficient approach to account for dispersion interactions is via an empirical Lennard-Jones (LJ) potential which is optimized for condensed phase and actively employed in classical force fields (similar approaches can be found elsewhere [65–67]). Despite the successes of adding a LJ dispersion term to the QM energy on small molecules [68], no rigorous tests have been performed in the literature on real macromolecules. Our QM MD simulations [69] with full LJ potential resulted in structure collapse. This suggests that the classical LJ contribution has to be scaled down, in agreement with others [67]. Indeed, part of the dispersion contribution is already incorporated into semiempirical Hamiltonians via parametrization. In the present QM calculation, the lack of dispersion interactions in the PM3 Hamiltonian was compensated by adding 50% of LJ interaction energy computed using CHARMM22 parameters [70]. This improved the ability of the model to better account for dispersion interactions while avoiding double counting of close-range dispersion interactions which are already partially included in the semiempirical methods. The LJ interactions were applied to all atoms of the BRCA1(BRCT) complexes and their constituents including the step of geometry optimization, and switched on the interval from 8 to 10 Å using CHARMM switching function [71].

Results and discussion

The P7 pentapeptide was synthesized and its binding free energy measured by ITC (cf. "Materials and methods" section, and Table 1). We observed a modest increase ($\Delta\Delta G = 0.3$ kcal/mol) in the binding affinity of P7 when compared to T1. This result is consistent with our hypothesis that the Nap side chain occupies a hydrophobic cluster on the N-terminus of the pSXXF binding site and results in a slightly tighter binding ligand.

Molecular dynamics simulations were performed on the series T1–T5 and P7 phosphopeptides complexed with BRCA1(BRCT). The root mean square difference (RMSD) of the protein backbone non-hydrogen atoms compared to the crystallographic geometry of the BRCA1(BRCT) protein is displayed in Fig. 3. BRCA1(BRCT) in complex with pSPVF (T5) showed the lowest RMSD of 1.2 Å among the series. The protein in other complexes is characterized with average RMSD of 1.5 Å. To test the feasibility of the replacement of pS by another negatively charged group, we tried the commonly used Glu substitution of pS in ligand T4. Unlike other studied peptides, T4 peptide did not form a stable complex with BRCA1(BRCT). After a few nanoseconds, T4 started separating from BRCA1(BRCT) and eventually left it. It has been recently found experimentally that T4 does not bind BRCA1(BRCT), and thus our MD simulation is consistent with the experimental data [9]. This result confirms that pS plays a key role in high affinity peptide binding to BRCA1(BRCT).

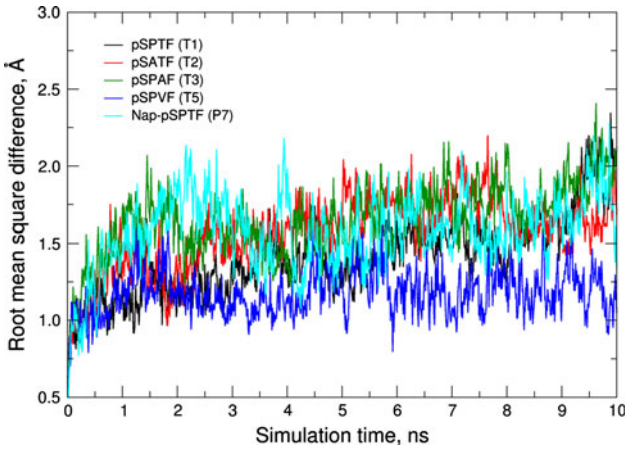

Fig. 3 RMSD of protein backbone atoms of BRCA1(BRCT) complexes along MD simulations using the crystallographic structure as a reference

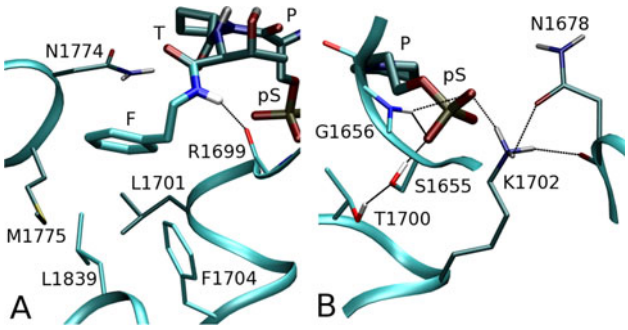

Fig. 4 T1-BRCA1(BRCT) interactions. Peptide T1 and receptor amino acids involved in the interaction are shown in licorice mode; distant amino acids are displayed in *darker color*. Color code for T1: H, *white*; C, *green*; N, *blue*; O, *red*. **A** T1(F)-BRCA1(BRCT) interactions. The receptor hydrophobic pocket is formed by L1701, F1704, N1774, M1775, and L1839. The peptide hydrogen of F makes a hydrogen bond with the carbonyl oxygen of R1699. **B** T1(pS)–BRCA1(BRCT) interactions via hydrogen-bonds OP(pS)···O_γ(S1655), OP(pS)···N(G1656), OP(pS)···N_z(K1702). The hydrophylic pocket of the receptor is additionally preconditioned for these interactions by hydrogen bonds S1655···T1700 and N1678···K1702. Figure was generated with VMD [72] and Pov-Ray

Characterization of Ac-pSPTF-CONH (T1) binding to BRCA1(BRCT)

During the MD simulations of T1 bound to BRCA1(BRCT), it was observed that the pS group of T1 forms a salt bridge with the side chain of K1702, and two hydrogen bonds with the side chain of S1655 and the peptide hydrogen of G1656 in the hydrophilic pocket of the receptor. These three interactions are stable throughout the MD simulation with average interatomic distances OP(pS)···N_z(K1702) = 2.8 ± 0.1 Å, OP(pS)···O_γ(S1655) = 2.6 ± 0.1 Å, and OP(pS)···N(G1656) = 3.1 ± 0.2 Å The pattern of hydrogen bonds established between pS and the receptor is illustrated on Fig. 4. It was also observed that these interactions were preserved among the other studied peptides (T2, T3, T5, P7). The MD simulations show no direct interaction between T1 and the side-chain of T1700, which contributes to delineate the pS pocket.

A second point of contact between T1 and BRCA1(BRCT) is at the P+3 (F) position (Fig. 4A). The phenyl ring of Phe is anchored in a hydrophobic pocket formed by amino acids L1701, F1704, N1774, M1775, and L1839. There is also a hydrogen bond between the backbone HN of Phe and the carbonyl oxygen of R1699, characterized by N(F)···O(R1699) = 3.0 ± 0.2 Å.

The above described interactions of pS and P+3 with the receptor are qualitatively observed in the trajectories of all studied complexes, and constitute the primary mode of peptide-BRCA1(BRCT) interaction. These data are consistent with previous experimental finding of primary role of pS and Phe in binding to BRCA1(BRCT) [9]. From the tight fit of Phe(P+3) side-chain within the hydrophobic pocket of the receptor, maintained throughout the simulation (Fig. 3), it is apparent that this pocket would not tolerate larger side-chains (cf. Fig. 4), in agreement with recent experimental data [18].

The role of mutation at the P+2 position in protein–ligand binding

The role of Thr(P+2) in T1 (pSPTF) was studied by mutating Thr to Ala (T3), and to valine (T5). According to isothermal calorimetric data, the Ala mutation caused a loss of 1.6 kcal/mol in binding affinity, whereas the Val mutation slightly enhanced binding by 0.3 kcal/mol [9]. These data confirm the important role that the P+2 position plays in protein–ligand binding. To show the impact of mutations at the P+2 site, the Thr → Ala mutation (T3) is analyzed relative to T1. The interaction of pS with the receptor remains unchanged; in the T3 complex, the average interatomic distances between phosphate oxygen atoms and the receptor amino acids are practically the same as in T1 complex (data not shown). Indeed, considering the strength of the electrostatic interaction between the charged phosphate group with three polar amino acids of the receptor it is hardly reasonable to expect significant perturbations at this site (Fig. 4). However, the analysis of the averaged trajectories of peptides T1, T3, and T5 bound to BRCA1(BRCT) shows important structural differences (Fig. 5).

It is readily seen that backbone atoms from pS to the P+2 residue show nearly a perfect structural alignment. Changes in the bound conformation start showing up right after the mutation at the P+2 position (Fig. 5). In the Thr → Ala mutation in T3, Phe bends from its position in T1. This could happen due to Ala offering less steric

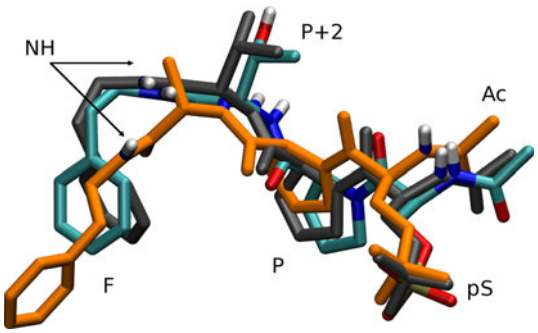

Fig. 5 Superimposed averaged structures of the bound conformation of T1, T3, and T5. In T1, H is *white*, C is *green*, N is *blue*, O is *red*; T3 is in *orange*, and T5 is in *gray*. P+2 is Thr, Ala, and Val in T1, T3 and T5, respectively. The alignment was performed using the backbone atoms of residues R1649–R1751, which represent half of the receptor size, and include the hydrophilic pocket of BRCA1(BRCT) which strongly interacts with pS, and which is structurally unaffected by the mutation at the P+2 position. Figure was generated with VMD [72] and Pov-Ray

hindrance to dihedral rotation at the P+2 position than that of Thr (T1) or Val (T5). Moreover, a Monte Carlo-based conformational search on free ligands T1 and T3 using generalized-Born implicit solvent model showed that T3 had 50% more accessible rotameric states than T1 or T5 (see Refs. [51–53] for computational details). This shows that T3 ligand is intrinsically more flexible than T1. Thus, T3 can interact with the receptor in different ways than T1. The superposition of complexes based on protein backbone atoms of amino acids R1649–R1751 is shown in Fig. 6.

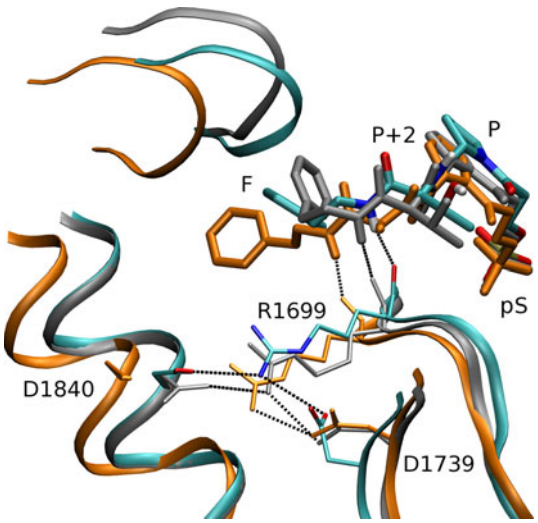

Fig. 6 Superimposed averaged complexes of T1, T3, and T5 with BRCA1(BRCT). In T1, H is *white*, C is *green*, N is *blue*, O is *red*; T3 is in *orange*, and T5 is in *gray*. In T1 and T5, the side chain of R1699 makes salt-bridge interactions with D1840 and D1739. In T3, the hydrogen bond with D1840 is lost. The amide hydrogen at the P+3 position (F) makes a hydrogen bond with the carbonyl oxygen of R1699. Figure was generated with VMD [72] and Pov-Ray

Compared to T1 and T5, the Ala substitution in T3 increases the average distance between the N of Phe(P+3) and the C_α of E1698 from 6.9 ± 0.3 and 6.7 ± 0.5 Å in T1 and T5, respectively, to 7.5 ± 0.4 Å in T3 (see Fig. 6). Thus, the hydrogen bond between the backbone HN of Phe and the carbonyl O of R1699 is geometrically shifted in T3 compared to T1 and T5 (see Fig. 6). In T1 and T5, the side chain of R1699 establishes salt-bridges with the side chains of D1840 and D1739, which are stable throughout the simulation (average distances in T1 are NH(R1699)···O_δ(D1840) = 1.9 ± 0.3 Å, NH(R1699)···O_δ(D1739) = 1.9 ± 0.1 Å, and 1.9 ± 0.4 and 2.0 ± 0.2 Å in T5, respectively). In T3, the salt bridge between R1699 and D1840 is disrupted, and only the salt bridge between R1699 and D1739 is retained (average distance NH(R1699)···O_δ (D1739) = 1.9 ± 0.2 Å). Another significant difference in T3 compared to T1 and T5 occurs at the P+3 position. In T1 and T5, the angle $\chi_{N-C\alpha-C\beta-C\gamma}$ is 72 ± 9° and 60 ± 10°, respectively. In T3, this angle is 169 ± 12° (Fig. 6), meaning that the Phe side chain is oriented quite differently than that in T1 and T5.

To expand our analysis, we considered the two arms of the receptor characterized by residues R1753–P1859 (right arm, which contains the hydrophilic pocket interacting with pS), and R1649–R1751 (left arm, which contains the hydrophobic pocket interacting with Phe, P+3) connected by a hinge at A1752. The relative position of the arms can be characterized by the angles defined by the C_α atoms of E1698–A1752–E1836 and S1655–A1752–N1774 (see Fig. 7). When T1 is bound, these angles are 49 ± 2° and 37 ± 1°, respectively, with similar angles for T5 (47 ± 2° and 41 ± 2°, respectively). In T3, however, the values are 55 ± 3° and 47 ± 3°, respectively. This means that in T3 the arms moved apart relative to their position in T1 and T5. Due to this opening, the average distance between the C_α atoms of E1698···E1836 and S1655···N1774 in T3 increased by 1.5 and 3.0 Å, respectively, compared to those in T1. If the backbone atomic coordinates of the left and right arms were treated separately, they would align nearly perfectly for T1, T3, and T5 complexes, what supports our division of BRCA1(BRCT) in two quasi-rigid arms.

Two factors could be identified related to the widening upon T3 binding. The Ala mutation at the P+2 position in T3 induces a change in the Phe(P+3) side-chain orientation, and a disruption of the salt-bridge between the side-chains of R1699···D1840. The new orientation of Phe requires more room to accommodate the ligand conformation, while the broken salt-bridge no longer contributes to holding arms together. Both these factors may originate in the increased flexibility of T3 compared to T1 and T5. It is not possible, however, to state which of these two factors is dominant. Additional studies will be necessary to decipher the mechanism of receptor widening.

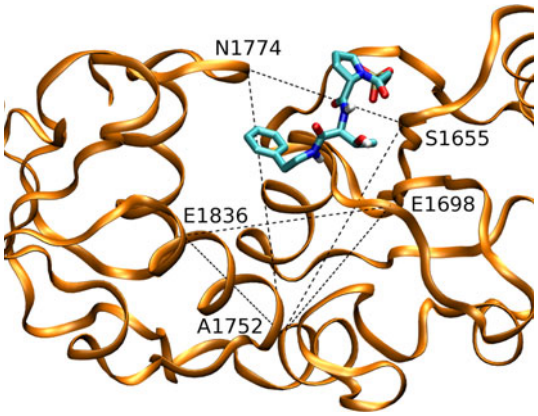

Fig. 7 Representation of the T1–BRCA1(BRCT) complex as consisting of *left* and *right* arms connected by a hinge around A1752. T1 is shown in licorice mode; H is *white*, C is *green*, N is *blue*, O is *red*. Upon T1 or T5 binding, the two arms stay closer than when T3 binds. Figure was generated with VMD [72] and Pov-Ray

The residual flexibility of T3 is further propagated to increase the rotational freedom of the NMe group, which terminates the peptide at the P+3 position. According to MD data, the standard deviation of the NMe group torsion angle as a measure of fluctuation is 17° in the T1 complex, whereas it increases to 50° in the T3 complex.

Our MD simulations confirmed that the Thr side-chain indeed makes no contacts with the receptor, as was previously postulated based on the analysis of X-ray structure of complex of BACH1 dodecapeptide with BRCA1 (BRCT) [9]. Indeed, the average distance O_γ(T)···O_ε (E1698) for T1 is 4.3 ± 1.0 Å. Moreover, we could not observe any intra-molecular hydrogen bonds involving Thr at P+2 in T1, either O_γ(T)···H(P + 4) or H_γ(T)···O(pS), what indirectly supports the relative insensitivity to the Thr → Val mutation (see below).

The role of Thr in peptide binding to BRCA1(BRCT) was additionally studied via the Thr → Val mutation (T5). The isothermal calorimetric study showed practically the same binding affinity for T5 as that of T1. Our structural analysis of the T1 and T5 complexes also shows overwhelming similarity between these two systems, while the Monte Carlo conformational search exhibited very similar number of accessible rotamers compared to T1. Therefore we conclude that this mutation makes practically no difference over T1. These results are in agreement with the experimental studies showing a minor role of the hydroxyl group of Thr on binding affinity [18].

The role of proline at the P+1 site

We also studied the Pro → Ala mutation at the P+1 position (T2). The complex of T2 with BRCA1(BRCT) shows the same structural characteristics discussed above for T1. This is consistent with the experimentally observed minimal impact that this mutation has on binding free energy. The observation that the Pro → Ala mutation (T2) brings less structural changes in the complex than the Thr → Ala one (T3), agrees with the previously made conclusion about the minor role of Pro in the ligand structure [9]. The Pro → Ala makes the ligand intrinsically more mobile which may explain the slight loss in the binding affinity of T2 over T1. This is consistent with an earlier conclusion based on the analysis of experimental data that Pro(P+1) and Thr(P+2) serve primarily as spacers between pS and Phe, thus providing optimal peptide backbone conformation to maximize the interaction of pS and Phe with the receptor [13]. The results of our binding free energy calculation for T1, T2, and T3 (Table 2) are also in agreement with experimental data [13] showing that the loss in binding affinity due to mutations at P+1 and P+2 positions is accompanied with an increase of entropy compared to T1.

Probing the hydrophobic pocket near the phosphate binding site

In earlier docking studies on BRCA1(BRCT) it was hypothesized that the BRCA1(BRCT) hydrophobic site constituted by V1654, L1657, and F1662 could be used to enhance protein–ligand binding affinity, by designing a ligand occupying this pocket [9]. To explore this possibility we extended the ligand at pS by adding a hydrophobic non-standard amino acid containing a naphthyl side chain (pentapeptide P7). The average MD structure of the P7 complex shows that the naphthyl group cannot reach that cleft due to geometric constraints stemming from the ligand anchored in two places [by the pS phosphate, and by the Phe(P+3)]. Forcing the naphthyl fragment to stay in that hydrophobic pocket would require displacing the phosphate group away from the electrophilic pocket, which is energetically unfavorable. Therefore, the naphthyl fragment adopts a conformation in which it makes favorable dispersion interactions with L1657, T1658, and P1659 of BRCA1(BRCT), what could explain some improvement in the binding affinity of P7 over that of T1. A longer and more flexible side chain than naphthyl would be necessary to reach the hydrophobic pocket lined up by V1654, L1657, and F1662 (cf. Fig. 8).

Protein–ligand binding free energy calculation

Computational data on binding free energy are presented in Table 2. MM/PB-SA predicts too favorable energies in the range of −30 kcal/mol, which exceeds the range of experimental values by an average factor of four. To estimate the quality of the prediction of relative energies we

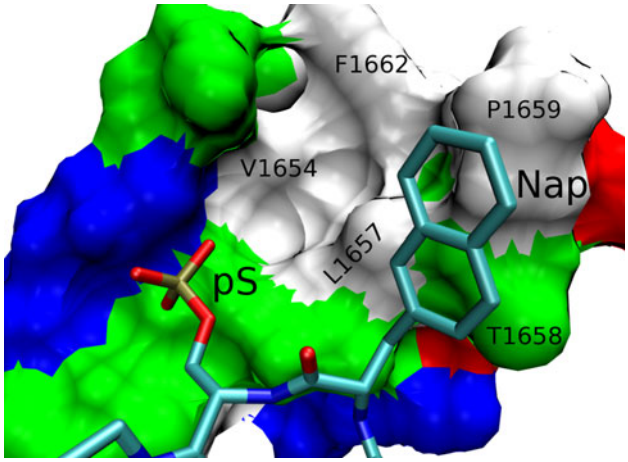

Fig. 8 Averaged structure of P7–BRCA1(BRCT) complex showing the solvent accessible surface of the receptor, and P7 in licorice mode. The positively charged amino acids are shown in *blue* on the surface, negatively charged in *red*, polar in *green*, and hydrophobic in *white*. In P7, C is *green*, N is *blue*, O is *red*. Geometrical restraints of the ligand backbone prevent the Nap side chain to reach the hydrophobic pocket defined by V1654, L1657, and F1662. Nap side-chain makes hydrophobic interactions with T1658 and P1659. Figure was generated with VMD [72] and Pov-Ray

employed Pearlman's predictive index (PI) [73]. This index assumes continuous values in the [−1, 1] range, where a value of −1 corresponds to systematically wrong predictions, zero corresponds to random, and a value of +1 corresponds to the ideal agreement in relative binding free energies between computational and experimental data. The PI associated with MM/PB-SA calculations is 0.71, representing a satisfactory performance.

The tendency to overestimate binding affinities is a known deficiency of the MM/PB-SA method [26, 74]. In an attempt to improve its predictability, Swanson et al. [46] performed an optimization of the atomic radii on a series of 14 polyalanine peptides using computed hydration free energies from TI simulations, given the absence of direct experimental data for such complex systems. These atomic radii were used in the present MM/PB-SA calculations showing a reasonable outcome in terms of relative binding free energy.

An alternative approach to improve the agreement between computational and experimental data was suggested by Purisima and co-workers [26, 54], scaling down the MM/PB-SA computed values, resulting in the empirical scoring function SIE. The corresponding binding free energies are presented in Table 2. There is an improvement in the absolute values of the computed binding free energies in closer agreement with experimental data. SIE predictions are characterized by a Pearlman index of 0.78. On the technical side, SIE calculation involves exactly the same steps as the standard MM/PB-SA calculation, which is evident from Eq. 8. Given its better accuracy, the SIE method may serve as a practical alternative to the standard MM/PB-SA method in computer-aided drug design [54]. However, a limiting factor of the SIE method is its explicit reliance on enthalpy-entropy compensation, which may not be complete. Besides, scaling down the Poisson–Boltzmann reaction field by a factor of ~ 10 ($\alpha = 0.104758$ in Eq. 8) masks the potential deficiencies of the used electrostatic model. This is why it is important to search for first-principle methods which offer a better computational efficiency and a wider range of applicability than FEP and TI.

One alternative approach is the QM description of the electrostatic interactions in protein–ligand systems. With

Table 2 Calculated protein–ligand binding free energies and their components

Ligand	$T\Delta S$[a]	ΔH (PB)[b]	ΔH (QM)[c]	ΔG (PB)[d]	ΔG (SIE)[e]	ΔG (QM)[f]	ΔG (Exp)
T1	−9.0 (0.2)	−40.7 (0.2)	−15.9 (0.2)	−31.7 (0.3)	−7.8 (0.01)	−6.9 (0.3)	−7.5 (0.2)
T2	−9.5 (0.4)	−38.0 (0.3)	−12.6 (0.2)	−28.4 (0.5)	−7.3 (0.02)	−3.1 (0.5)	−6.7 (0.3)
T3	−10.1 (0.2)	−37.2 (0.2)	−13.7 (0.2)	−27.0 (0.3)	−7.4 (0.01)	−3.6 (0.3)	−5.9 (0.2)
T4	No binding						
T5	−8.8 (0.4)	−36.4 (0.2)	−16.5 (0.2)	−27.6 (0.4)	−7.5 (0.01)	−7.7 (0.5)	−7.8 (0.1)
P7	−12.9 (0.2)	−47.8 (0.2)	−17.3 (0.3)	−34.9 (0.2)	−8.7 (0.01)	−4.4 (0.3)	−7.8 (0.1)
PI[g]	NA	NA	NA	0.71	0.78	0.78	NA

Values of energies in kcal/mol

The values in brackets represent the error of the mean

[a] Entropic contribution calculated according to Eq. 4

[b] Enthalpic part from MM/PB-SA calculation including internal and solvation free energy

[c] Enthalpic part from MM-QMSA calculation including internal energy and solvation free energy

[d] Binding free energy from MM/PB-SA calculation (Eq. 1)

[e] Binding free energy from SIE calculation (Eq. 8)

[f] Binding free energy from MM-QMSA calculation (Eq. 9)

[g] Pearlman's predictive index

the advent of modern computer hardware it is technically feasible to perform MM/PB-SA-like calculations on protein–ligand complexes using QM and a continuum solvent model. Such calculaions were previously reported by us on a series of phosphopeptide inhibitors of p56 LCK SH2 domain using 1,000 snapshots generated by MD with AMBER force field, followed by geometry optimization of the snapshots with the semiempirical PM3 Hamiltonian coupled with the implicit solvent model COSMO [27]. In this work we applied a similar strategy to calculate the binding free energy of peptides to the BRCA1(BRCT) system using QM.

The results of the MM-QMSA binding free energy calculations on BRCA1(BRCT) complexes presented in Table 2, show good agreement with experimental data. Similar to SIE, the QM predictions are characterized by an improved Pearlman index of 0.78 relative to the standard MM/PB-SA calculation (Pearlman index of 0.71). However, unlike SIE, the improvement in accuracy of prediction is obtained from a first-principle calculation method. This is a significant result because no other first-principle method (FEP or TI) may be applied to BRCA1(BRCT) and phosphopeptide ligands due to large net charge carried by the protein and ligands. The correlation coefficients for MM/PB-SA, SIE and MM-QMSA are 0.59, 0.61 and 0.69, respectively. Although the number of points is small to draw conclusions with statistical meaning, it shows an improvement in going towards QM. In this work, the LJ contribution scaled by 50% was found to provide the best performance across the set of peptides. However, the difference in calculated and experimental binding free energies, especially for T2 and P7 ($\sim$3 kcal/mol), shows that additional improvement could be achieved by improving the balance between the QM term and the dispersion contribution.

First-principle calculation of protein–ligand binding affinity is particularly challenging in the case of phosphopeptides binding to BRCA1(BRCT) due to the strong electrostatic interactions; simulations in such systems might require accounting for the effects of electronic polarization in protein–ligand and solute–solvent interactions, what is directly accounted for in the QM framework due to the treatment of electrostatic interactions via charge density which self-consistently adjusts to the electric field of the environment. Moreover, it is well-known that the description of biological systems can be improved by accounting for polarization effects [75, 76].

In fact, the success of a QM treatment of electrostatic interactions as part of the MM/PB-SA-like computational protocol also suggests that the electrostatic properties of protein and ligand in the free state are different from their state in the protein–ligand complex. This could explain the sensitivity of MM/PB-SA to insufficient charge transferability, displaying in some cases an improper balance between the intermolecular electrostatic interaction and the electrostatic contribution of the solvent calculated via a continuum solvent model. The importance of ad hoc derived charges has been already shown [77].

While the application of polarizable force fields in binding free-energy calculation is still an area of ongoing development, it is already feasible to account for electrostatic polarization in protein–ligand complexes by means of semiempirical QM methods.

Conclusions

The computational study of inhibitor binding to BRCA1(BRCT) confirms that the studied peptides are anchored by a two-point interaction involving the phosphate group of pS and the Phe residue at the P+3 position. It was observed that mutation at the P+2 position controls structural flexibility of the peptide, which is less affected by mutations at the P+1 position. We also observed that A1752 could be considered a hinge between the left and right arms of the receptor: upon binding of peptides T1 and T5, which are characterized by a lesser backbone flexibility, these arms stay closer together. The Thr → Ala mutation at the P+2 position in T3 leads to receptor opening. No intra-peptide hydrogen bond was observed throughout the simulation involving Thr(P+2) in T1. Thus, it is not surprising that the Thr → Val mutation (T5) had very similar experimental binding energy, and that the simulation of T5 bound to BRCA1(BRCT) exhibited a very similar behavior to that of T1. From ITC measurements, it was determined that P7 displayed a modest increase in binding free energy compared to T1, and the dispersion interactions of the Nap side chain of P7 within the hydrophobic patch observed during the MD simulations correlate well with experiments. The pSPTF and pSVTF signatures exhibited the highest affinities, and showed conserved interactions with the protein during the simulations. This, coupled with the fact that to reach the hydrophobic pocket lined up by V1654, L1657, and F1662 (cf. Fig. 8) a longer side chain than naphthyl would be necessary, led to the synthesis of a small library of peptide mimics with varied hydrophobic groups and longer linkers attached to the N-terminus of the phosphoserine residue on pSPTF and pSVTF scaffolds. Evaluation of this library may lead to the identification of a high affinity BRCT(BRCA1) binding peptide mimic (in progress).

Calculations with the MM/PB-SA protocol using optimized atomic radii resulted in relative binding free energy calculations showed a reasonable agreement with experimental results, though it resulted in too favorable absolute binding free energies by an average factor of four. The SIE method exhibited a better performance, though its empirical

character poses uncertainties regarding its general applicability. The MM-QMSA method [27], a QM-based end-point method in which the fixed-charge electrostatic model is replaced by a QM treatment of electrostatic interactions using the semiempirical PM3 Hamiltonian, and the continuum solvent model COSMO, predicted binding free energies in better agreement with experimental values. The insufficiently favorable binding free energy prediction for P7 in QM calculation suggests there is room for additional improvement in reaching a better balance between the QM calculation and the treatment of dispersion interactions via a LJ term, for example, by integrating the dispersion interactions into semiempirical parameters.

The present study demonstrates that too favorable binding affinity predictions of the classical MM/PB-SA method persist despite using an optimized set of atomic radii directly optimized to reproduce hydration free energies of biomolecules using the Poisson–Boltzmann implicit solvent model. This may also suggest that when using a continuum solvent model in the classical mechanics approach, it remains questionable whether fixed charges would provide equally good representation of the bound and free states of the biomolecules, since ligand and protein would experience very different environments in each case, and different sets of charges would be necessary to account for this effect. This balance appears to be better accounted for by the QM treatment of electrostatic interactions in the MM-QMSA method, what provides further motivation to continue developing of QM-based first-principle end-point methods.

In summary, the interactions of the C-terminus domains of BRCA1(BRCT) protein with phosphopeptide inhibitors have been studied by experimental and computational methods in order to characterize those interactions at an atomic level. As a complement, binding free energy was evaluated by three different methods, namely MM/PB-SA, SIE and the QM-based MM-QMSA. Although further systematic validation on other systems is necessary, it is encouraging that QM binding free energy calculations can be performed with an adequate balance of accuracy and computational efficiency.

Acknowledgments The authors acknowledge the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing high-performance computing resources (http://www.tacc.utexas.edu). This work was supported by the R. A. Welch Foundation Chemistry and Biology Collaborative Grant from the John S. Dunn Gulf Coast Consortium for Chemical Genomics (to CNC and AN) and NIH (R01CA127239 to AN).

References

1. Koonin EV, Altschul SF, Bork P (1996) BRCA1 protein products: functional motifs. Nat Genet 13(3):266–268. doi:10.1038/ng0796-266
2. Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, Livingston DM (2001) BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell 105(1):149–160
3. Yu X, Chen J (2004) DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. Mol Cell Biol 24(21):9478–9486. doi:10.1128/MCB.24.21.9478-9486.2004
4. Wang B, Matsuoka S, Ballif BA, Zhang D, Smogorzewska A, Gygi SP, Elledge SJ (2007) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science 316(5828):1194–1198. doi:10.1126/science.1139476
5. Manke IA, Lowery DM, Nguyen A, Yaffe MB (2003) BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science 302(5645):636–639. doi:10.1126/science.1088877
6. Yu X, Chini CC, He M, Mer G, Chen J (2003) The BRCT domain is a phospho-protein binding domain. Science 302(5645):639–642. doi:10.1126/science.1088753
7. Kim H, Huang J, Chen J (2007) CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. Nat Struct Mol Biol 14(8):710–715. doi:10.1038/nsmb1277
8. Kennedy RD, Quinn JE, Mullan PB, Johnston PG, Harkin DP (2004) The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst 96(22):1659–1668. doi:10.1093/jnci/djh312
9. Lokesh GL, Muralidhara BK, Negi SS, Natarajan A (2007) Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design. J Am Chem Soc 129(35):10658–10659. doi:10.1021/ja0739178
10. Botuyan MV, Nomine Y, Yu X, Juranic N, Macura S, Chen J, Mer G (2004) Structural basis of BACH1 phosphopeptide recognition by BRCA1 tandem BRCT domains. Structure 12(7):1137–1146. doi:10.1016/j.str.2004.06.002
11. Campbell SJ, Edwards RA, Glover JN (2010) Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1. Structure 18(2):167–176. doi:10.1016/j.str.2009.12.008
12. Clapperton JA, Manke IA, Lowery DM, Ho T, Haire LF, Yaffe MB, Smerdon SJ (2004) Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. Nat Struct Mol Biol 11(6):512–518. doi:10.1038/nsmb775
13. Joseph PR, Yuan Z, Kumar EA, Lokesh GL, Kizhake S, Rajarathnam K, Natarajan A (2010) Structural characterization of BRCT-tetrapeptide binding interactions. Biochem Biophys Res Commun 393(2):207–210. doi:10.1016/j.bbrc.2010.01.098
14. Shen Y, Tong L (2008) Structural evidence for direct interactions between the BRCT domains of human BRCA1 and a phosphopeptide from human ACC1. Biochemistry 47(21):5767–5773. doi:10.1021/bi800314m
15. Shiozaki EN, Gu L, Yan N, Shi Y (2004) Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. Mol Cell 14(3):405–412
16. Varma AK, Brown RS, Birrane G, Ladias JA (2005) Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex. Biochemistry 44(33):10941–10946. doi:10.1021/bi0509651
17. Williams RS, Lee MS, Hau DD, Glover JN (2004) Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nat Struct Mol Biol 11(6):519–525. doi:10.1038/nsmb776
18. Yuan Z, Kumar EA, Kizhake S, Natarajan A (2011) Structure–activity relationship studies to probe the phosphoprotein binding site on the carboxy terminal domains of the breast cancer susceptibility gene 1. J Med Chem. doi:10.1021/jm1016413

19. Leung CC, Gong Z, Chen J, Glover JN (2011) Molecular basis of BACH1/FANCJ recognition by TopBP1 in DNA replication checkpoint control. J Biol Chem 286(6):4292–4301. doi:10.1074/jbc.M110.189555
20. Cavasotto CN, Singh N (2008) Docking and high throughput docking: successes and the challenge of protein flexibility. Curr Comput Aided Drug Design 4:221–234
21. Spyrakis F, Bidon-Chanal A, Barril X, Luque FJ (2011) Protein flexibility and ligand recognition: challenges for molecular modeling. Curr Top Med Chem 11:192–210
22. Straatsma TP, McCammon JA (1991) Theoretical calculations of relative affinities of binding. Methods Enzymol 202:497–511
23. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492. doi:10.1146/annurev.bb.18.060189.002243
24. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J Am Chem Soc 120(37):9401–9409
25. Vorobjev YN, Hermans J (1999) ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophys Chem 78(1–2):195–205
26. Naïm M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO (2007) Solvated interaction energy (SIE) for scoring protein–ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model 47(1):122–133. doi:10.1021/ci600406v
27. Anisimov VM, Cavasotto CN (2011) Quantum mechanical binding free-energy calculation for phosphopeptide inhibitors of the Lck SH2 domain. J Comput Chem 32:2254–2263. doi:10.1002/jcc.21808
28. Fanfrlik J, Bronowska AK, Rezac J, Prenosil O, Konvalinka J, Hobza P (2010) A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. J Phys Chem B 114(39):12666–12678. doi:10.1021/jp1032965
29. Gräter F, Schwarzl SM, Dejaegere A, Fischer S, Smith JC (2005) Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. J Phys Chem B 109(20):10474–10483. doi:10.1021/jp044185y
30. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM Jr (2007) The role of quantum mechanics in structure-based drug design. Drug Discov Today 12(17–18):725–731
31. Soderhjelm P, Aquilante F, Ryde U (2009) Calculation of protein–ligand interaction energies by a fragmentation approach combining high-level quantum chemistry with classical many-body effects. J Phys Chem B 113(32):11085–11094. doi:10.1021/jp810551h
32. Zhou T, Huang D, Caflisch A (2008) Is quantum mechanics necessary for predicting binding free energy? J Med Chem 51(14):4280–4288. doi:10.1021/jm800242q
33. Friesner RA, Guallar V (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. Annu Rev Phys Chem 56:389–427
34. Illingworth CJR, Morris GM, Parkes KEB, Snell CR, Reynolds CA (2008) Assessing the role of polarization in docking. J Phys Chem A 112(47):12157–12163. doi:10.1021/jp710169m
35. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65(3):712–725. doi:10.1002/prot.21123
36. Dupradeau F-Y, Cézard C, Lelong R, Stanislawiak É, Pêcher J, Delepine JC, Cieplak P (2008) R.E.DD.B.: a database for RESP and ESP atomic charges, and force field libraries. Nucleic Acids Res 36(suppl 1):D360–D367. doi:10.1093/nar/gkm887
37. Homeyer N, Horn A, Lanig H, Sticht H (2006) AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. J Mol Model 12(3):281–289
38. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802
39. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
40. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341
41. Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER9. University of California, San Francisco
42. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii. J Phys Chem A 101(16):3005–3014
43. Feig M, Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 14(2):217
44. Wan S, Coveney PV, Flower DR (2005) Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, Poisson–Boltzmann and linear interaction energy approximations. Philos Trans A Math Phys Eng Sci 363(1833):2037–2053. doi:10.1098/rsta.2005.1627
45. Gilson MK, Zhou H-X (2007) Calculation of protein–ligand binding affinities. Annu Rev Biophys Biomol Struct 36(1):21–42. doi:10.1146/annurev.biophys.36.040306.132550
46. Swanson JMJ, Adcock SA, McCammon JA (2005) Optimized radii for Poisson–Boltzmann calculations with the AMBER force field. J Chem Theory Comput 1(3):484–493. doi:10.1021/ct049834o
47. Tsui V, Case DA (2000) Theory and applications of the generalized born solvation model in macromolecular simulations. Biopolymers 56(4):275–291. doi:10.1002/1097-0282(2000)56:4<275:aid-bip10024>3.0.co;2-e
48. Swanson JMJ, Henchman RH, McCammon JA (2004) Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 86(1):67–74
49. Lee MS, Olson MA (2006) Calculation of absolute protein–ligand binding affinity using path and endpoint approaches. Biophys J 90(3):864–877
50. Li Z, Scheraga HA (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci USA 84(19):6611–6615
51. Li W, Cavasotto CN, Cardozo T, Ha S, Dang T, Taneja SS, Logan SK, Garabedian MJ (2005) Androgen receptor mutations identified in prostate cancer and androgen insensitivity syndrome display aberrant ART-27 coactivator function. Mol Endocrinol 19(9):2273–2282
52. Monti MC, Casapullo A, Cavasotto CN, Napolitano A, Riccio R (2007) Scalaradial, a dialdehyde-containing marine metabolite

that causes an unexpected noncovalent PLA(2) inactivation. ChemBioChem 8(13):1585–1591
53. Torra IP, Ismaili N, Feig JE, Xu CF, Cavasotto C, Pancratov R, Rogatsky I, Neubert TA, Fisher EA, Garabedian MJ (2008) Phosphorylation of liver X receptor alpha selectively regulates target gene expression in macrophages. Mol Cell Biol 28(8): 2626–2636
54. Cui Q, Sulea T, Schrag JD, Munger C, Hung M-N, Naïm M, Cygler M, Purisima EO (2008) Molecular dynamics-solvated interaction energy studies of protein–protein interactions: the MP1-p14 scaffolding complex. J Mol Biol 379(4):787–802. doi: 10.1016/j.jmb.2008.04.035
55. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16):1999–2012. doi: 10.1002/jcc.10349
56. Lee MC, Duan Y (2004) Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. Proteins 55(3):620–634. doi:10.1002/prot.10470
57. Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2(5):799–805
58. Daniels AD, Scuseria GE, Farkas Ö, Schlegel HB (2000) Geometry optimization of Kringle 1 of plasminogen using the PM3 semiempirical method. Int J Quantum Chem 77(1):82–89. doi:10.1002/(sici)1097-461x(2000)77:1<82:aid-qua9>3.0.co;2-3
59. Stewart JJP (2009) Application of the PM6 method to modeling proteins. J Mol Model 15:765–805
60. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S (2005) A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. J Med Chem 48(17):5437–5447
61. Bugaenko VL, Bobrikov VV, Andreyev AM, Anikin NA, Anisimov VM (2009) LocalSCF 2.1. Fujitsu Ltd, Tokyo
62. Stewart JJP (1989) Optimization of parameters for semiempirical methods I. Method. J Comput Chem 10(2):209–220
63. Anisimov VM, Cavasotto CN (2011) Hydration free energies using semiempirical quantum mechanical Hamiltonians and a continuum solvent model with multiple atomic-type parameters. J Phys Chem B 115:7896–7905
64. Dewar MJS, Thiel W (1977) A semiempirical model for the two-center repulsion integrals in the NDDO approximation. Theor Chim Acta 46(2):89–104. doi:10.1007/bf00548085
65. Foster ME, Sohlberg K (2010) Empirically corrected DFT and semi-empirical methods for non-bonding interactions. Phys Chem Chem Phys 12(2):307–322
66. McNamara JP, Hillier IH (2007) Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. Phys Chem Chem Phys 9(19):2362–2370
67. Rezac J, Fanfrlik J, Salahub D, Hobza P (2009) Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. J Chem Theory Comput 5:1749–1760
68. He X, Fusti-Molnar L, Cui G, Merz KM (2009) Importance of dispersion and electron correlation in ab initio protein folding. J Phys Chem B 113(15):5290–5300. doi:10.1021/jp8106952
69. Anisimov VM, Bugaenko VL, Cavasotto CN (2009) Quantum mechanical dynamics of charge transfer in ubiquitin in aqueous solution. ChemPhysChem 10(18):3194–3196
70. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616
71. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614. doi:10.1002/jcc.21287
72. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graphics 14(1):33–38
73. Pearlman DA, Charifson PS (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. J Med Chem 44(21): 3417–3423
74. Woo HJ, Roux B (2005) Calculation of absolute protein–ligand binding free energy from computer simulations. Proc Natl Acad Sci USA 102(19):6825–6830
75. MacKerell AD Jr (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25(13): 1584–1604
76. Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85
77. Cho AE, Guallar V, Berne BJ, Friesner R (2005) Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. J Comput Chem 26(9):915–931