Journal of Molecular Graphics and Modelling 61 (2015) 281–289

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Folding and binding energy of a calmodulin-binding cell antiproliferative peptide

Ahmad M. Almudallal[a], Ivan Saika-Voivod[a,*], John M. Stewart[b]

[a] Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NL A1B 3X7, Canada
[b] Soricimed Biopharma Inc., Sackville, NB E4L 4H2, Canada

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form 31 July 2015
Accepted 3 August 2015
Available online 7 August 2015

Keywords:
Peptide modelling
Peptide binding
Docking
Drug design
Molecular dynamics simulation

ABSTRACT

We carry out a computational study of a calmodulin-binding peptide shown to be effective in reducing cell proliferation. We find several folded states for two short variants of different length of the peptide and determine the location of the binding site on calmodulin, the binding free energy for the different conformers and structural details that play a role in optimal binding. Binding to a hydrophobic pocket in calmodulin occurs via an anchoring phenylalanine residue of the natively disordered peptide, and is enhanced when a neighbouring hydrophobic residue acts as a co-anchor. The shorter sequence possesses better binding to calmodulin, which is encouraging in terms of the development of non-peptide analogues as therapeutic agents.

1. Introduction

Many disease states result from uncontrolled cell proliferation. Identifying agents that can safely and selectively inhibit proliferation pathways provides a basis for developing therapies against such conditions as cancer and restenosis. A recent patent [1] identifies a family of peptides that targets the interaction between proteins calmodulin (CaM) and cyclin E, and thus holds promise for treating unwanted proliferation in vascular smooth muscle cells and certain cancers.

While determination of the three-dimensional structure of proteins, so vital for rational drug design, has traditionally fallen to X-ray crystallography for crystallizing proteins and NMR for those that do not, e.g., membrane proteins [2,3], computational approaches, including both sequential and structural bioinformatics (see, e.g., [4,5] and a recent review [6]) complement these time-consuming and expensive techniques in the pursuit of drug development. Moreover, dynamic aspects of the functioning of proteins [7], such as cooperative effects [8], allosteric transitions [9] and intercalation of drugs into DNA [10], benefit particularly well from techniques like MD simulation. With these considerations in mind, we employ computational methods in the hope of uncovering information useful for developing therapies against cell-proliferation related diseases.

In this study, we turn to computer simulations in order to gain a better understanding of how these antiproliferative peptides interact with bovine CaM. In particular, we perform simulations of two peptides with sequences VTVFL (CBS-5), one of the shortest sequences identified as having potential therapeutic properties, and ANVTVFLQD (CBS-9), a slightly longer variant.

We use a complementary set of computer simulations: molecular dynamics simulations of peptide in water to quantify the tendency of the peptide to fold in addition to identifying folded states or conformers of the peptides; docking simulations to find the preferred attachment sites of the different conformers on CaM; and umbrella sampling simulations to determine the binding free energies of the conformers with CaM. By analyzing simulation data for the thermalized bound state, we identify conformational subtleties of the peptide that lead to better binding.

2. Methods

In this study we carry out molecular dynamics (MD) simulations and related analyses with version 4.5.5 of the GROMACS software suite [11–15] of CBS-5 and CBS-9, in water alone, as well as with bovine CaM [16], the structure of which is available through the Protein Data Bank [17,18]. We use the AMBER99SB [19] interaction potential along with the TIP4P-EW model of water [20]. The peptides are constructed with standard NH_3^+ and COO^- termini with

* Corresponding author.
E-mail address: saika@mun.ca (I. Saika-Voivod).

http://dx.doi.org/10.1016/j.jmgm.2015.08.002
1093-3263/

SwissPdb Viewer software [21]. VMD software is used to generate graphical representations of peptide and protein [22].

Simulations of peptide alone in water done to identify folded configurations are carried out at constant temperature $T = 37\,^{\circ}C$, kept fixed with the Nosé–Hoover algorithm with a time constant of 0.2 ps [23,24]. The pressure is held constant at $P = 1$ atm with the Parrinello–Rahman algorithm, employing a time constant of 5 ps and compressibility of 4.5×10^{-5} bar^{-1} [25,26]. One hundred independent trajectories of length 150 ns are each seeded with a configuration drawn every 1 ns from a constant volume simulation at 727 °C. There are 1387 and 1361 water molecules for these simulations of CBS-5 and CBS-9 peptides, respectively, and the box has average side length 3.47 nm. For CBS-9, we add one Na^+ counter ion to maintain charge neutrality.

We use a selection of compact conformations of the peptides, as well as an extended structure to act as a control, to search for favourable binding sites on CaM using ZDOCK (version 3.0.2) [27], which treats both peptide and CaM as rigid bodies as it efficiently searches many trial positions and orientations of the peptide to find favourable configurations between the two. ZDOCK quantifies the docked configurations with a zscore, the higher the better, based on a statistical potential derived from a database of known interactions, shape complementarity and electrostatics [28]. Although ZDOCK produces 2000 docked configurations for each peptide conformer, we only use the best one for each conformer for subsequent analysis.

We rationalize our focus on compact configurations based on three considerations: we anticipate that binding is driven by hydrophobic interactions (the peptides are hydrophobic) and folding tends to expose the hydrophobic side chains in our case; our peptides are derived from larger ones, which are likely folded in some way (CBS-9 already shows a tendency to fold); even in the case of intrinsically disordered peptides, folding occurs upon binding [29,30].

Binding free energies for CBS-5 and CBS-9 to CaM are calculated starting from the best docked position of each conformer using the MD-based "pull" code [31] and umbrella sampling (US) [32–34] capabilities of GROMACS. Using only the best docked configuration for each conformer is motivated in our case by the observation that only a single, dominant binding arrangement emerges for the conformers, and it is shared by all the top-ranked docking configurations. The pull code begins with a docked configuration from ZDOCK, and slowly pulls the peptide and calmodulin away from each other. System configurations sampled from a series of distances provide starting configurations for US simulations. Each US simulation is done with over 22,000 water molecules, and necessitates 26 (for CBS-9) and 25 (for CBS-5) simulation replicas. We add Na^+ counter ions to maintain charge neutrality (15 for CaM, 1 for CBS-9 and none for CBS-5). Each replica constrains the peptide, through a harmonic potential, to be a certain distance away from a specific atom of CaM located just behind the binding site, i.e., there is a potential energy added to the Hamiltonian of the system,

$$U_c = \kappa(r - r_0)^2, \tag{1}$$

where $\kappa = 1000$ kJ mol^{-1} nm^{-2} gives the strength of the constraint, r is the distance between the centre of mass of the peptide and the chosen CaM atom, and r_0 sets the target distance to be sampled. We run MD simulations at $T = 37\,^{\circ}C$ and $P = 1$ atm of each replica for 50 ns, reporting data from the second half of the simulations. Distributions (histograms) of CaM-peptide distances from each of

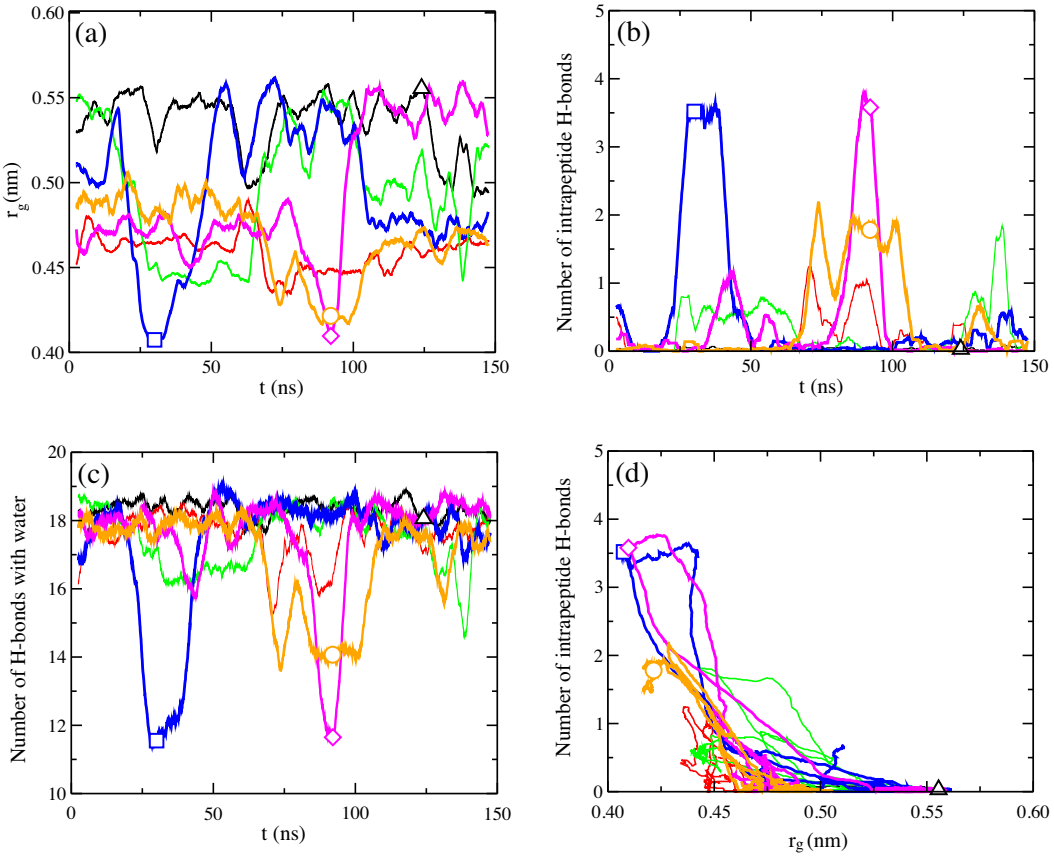

Fig. 1. Sampling of data from CBS-5 in water only at $T = 37\,^{\circ}C$. Panel (a) shows the r_g of the peptide as a function of time, (b) shows the number of H-bonds formed within the peptide itself as a function of time, (c) shows the number of H-bonds between the peptide and the surrounding water, while panel (d) shows intrapeptide H-bonds as a function of r_g. Symbols indicate simulation run and time from which conformers CBS-5a (square), CBS-5b (diamond), CBS-5c (circle) and CBS-5x (triangle) are taken. Data points are running averages over 5 ns.

the replicas are used to generate the free energy (reversible work) required to pull the peptide out of the binding site by calculating the potential of mean force (PMF). The values of r_0 are spaced 0.15 nm apart in adjacent replicas to ensure sufficient histogram overlap. Histogram analysis is carried out with the Weighted Histogram Analysis Method [35,36]. Finally, we analyze configurations from the lowest free energy states, i.e., the bound states, to identify CaM residues involved in the binding, and to determine configurational features of the peptide that enhance binding.

3. Results

3.1. Folded structure of the peptides

The peptide simulations in water alone yield an ensemble of folded conformations, starting points for determining how the peptide interacts with CaM. In Fig. 1(a) we plot a sample of time series of the gyration radius r_g from our 100 independent simulations at $T = 37\,^{\circ}$C of CBS-5. Fig. 1(b)–(d) shows that large decreases in r_g are accompanied by the formation of hydrogen bonds internal to the peptide and a commensurate decrease in the number of H-bonds between the peptide and water. Shown in this sampling are the three best cases in terms of having coincidentally the largest number of internal H-bonds and the smallest r_g. Identifying these different folded states reduces the effort required to find good bound states, while allowing us to discern structural details that lead to good binding.

The three folded configurations of CBS-5 chosen as starting points for further simulations are CBS-5a, taken from the blue curve in Fig. 1 near $t = 30$ ns (indicated by a square); CBS-5b, from the magenta curve near $t = 90$ ns (diamond); and CBS-5c, from the orange curve near $t = 90$ ns (circle). These three conformers are shown in Fig. 2. While they have different H-bonding, they all share roughly the same overall shape. CBS-5a and CBS-5b in particular are very similar. Shown in Fig. 2(d) is an extended version of the peptide, CBS-5x, which we also analyze (labelled by a triangle in Fig. 1).

We note that CBS-5 tends to be in an extended configuration in water. We quantify this by calculating the probability distribution $P_r(r_g)$ for r_g over all 100 MD runs. In Fig. 4(a), we plot the free energy derived from this probability,

$$\frac{\Delta G(r_g)}{k_B T} = -\ln[P_r(r_g)], \tag{2}$$

where k_B is Boltzmann's constant, and see that the folded structures are roughly 3 k_BT higher in free energy than extended conformations.

We carry out the same procedure for CBS-9, and choose four folded structures for subsequent analysis, labelled CBS-9a through CBS-9d and shown in Fig. 3. The folded structures for CBS-9 are similar to those of CBS-5, in that they are also non-helical and simply bent, roughly in the middle, with the two resulting arms zipped together with H-bonds. For both CBS-5 and CBS-9, the folding results in hydrophobic side chains being exposed to water. The difference is that CBS-9 possesses a greater propensity to fold, with extended and compact structures being roughly equally probable, as shown in Fig. 4(b).

In Table 1, we list the root mean square deviation (RMSD) between the positions of the backbone atoms of our conformers, as calculated by GROMACS utility g_confrms. Only atoms common to both CBS-5 and CBS-9 (the VTVFL portion) are considered. The values indicate that CBS-5a and CBS-5b are nearly the same, that CBS-5x is indeed quite different from the rest, and that otherwise, the conformers generally sample significantly different configurations.

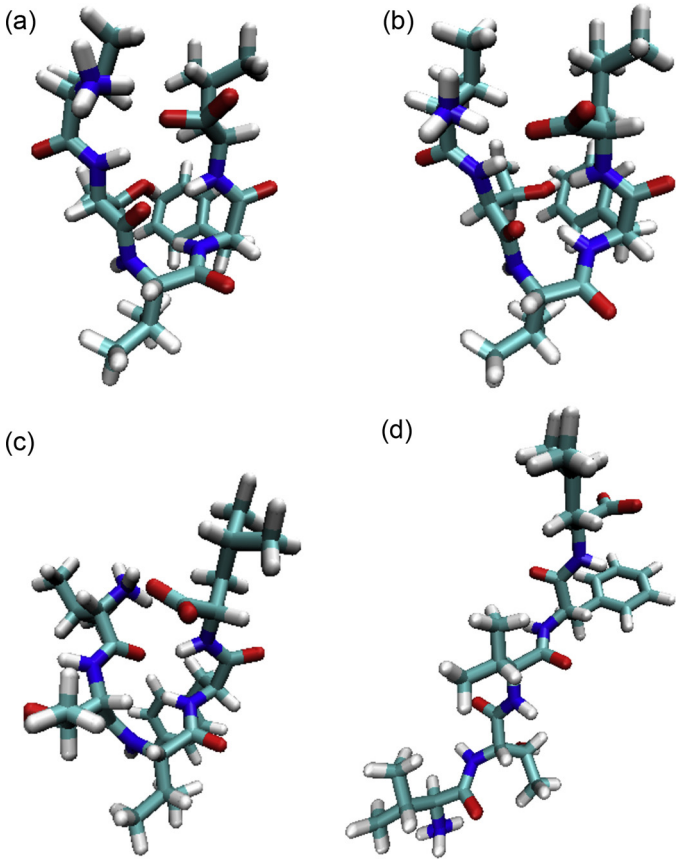

Fig. 2. Snapshots of CBS-5 taken from each of the dips in the r_g in Fig. 1. (a) CBS-5a, corresponding to the blue square. (b) CBS-5b, corresponding to the magenta diamond. (c) CBS-5c, corresponding to the orange circle. Panel (d) shows CBS-5x, a typical extended configuration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

3.2. Docking

Having obtained a few candidate structures for CBS-5 and CBS-9, we use ZDOCK to determine potential docking sites on CaM. An examination of the best-scoring docked configurations reveals a single, dominant binding arrangement. The best docked configuration from each of the eight conformers put the peptide in a position similar to the one shown in Fig. 5(a), in which the F residue anchors the peptide to a hydrophobic pocket near the N-terminal lobe of CaM, which is shown in Fig. 5(b). This hydrophobic docking site is different from at least two other sites identified computationally on CaM [37], but is involved with hydrophobic bonding of longer, helical peptides [38,39]. If we expand our view to the best five docked configurations for each of our eight conformers (to consider 40 docking arrangements), 30 (75%) of them have F as an anchor at this same binding site, 7 (17.5%) of them bind to the same site but with different hydrophobic residues as anchor (it varies as to which

Table 1
RMSD values (nm) for backbone atoms between all peptide conformers. Only atoms common to CBS-5 and CBS-9 are considered in the comparison.

	5b	5c	5x	9a	9b	9c	9d
5a	0.021	0.122	0.392	0.109	0.178	0.156	0.224
5b	–	0.123	0.381	0.100	0.174	0.150	0.218
5c	–	–	0.410	0.177	0.132	0.157	0.262
5x	–	–	–	0.314	0.319	0.320	0.266
9a	–	–	–	–	0.184	0.153	0.177
9b	–	–	–	–	–	0.139	0.217
9c	–	–	–	–	–	–	0.176

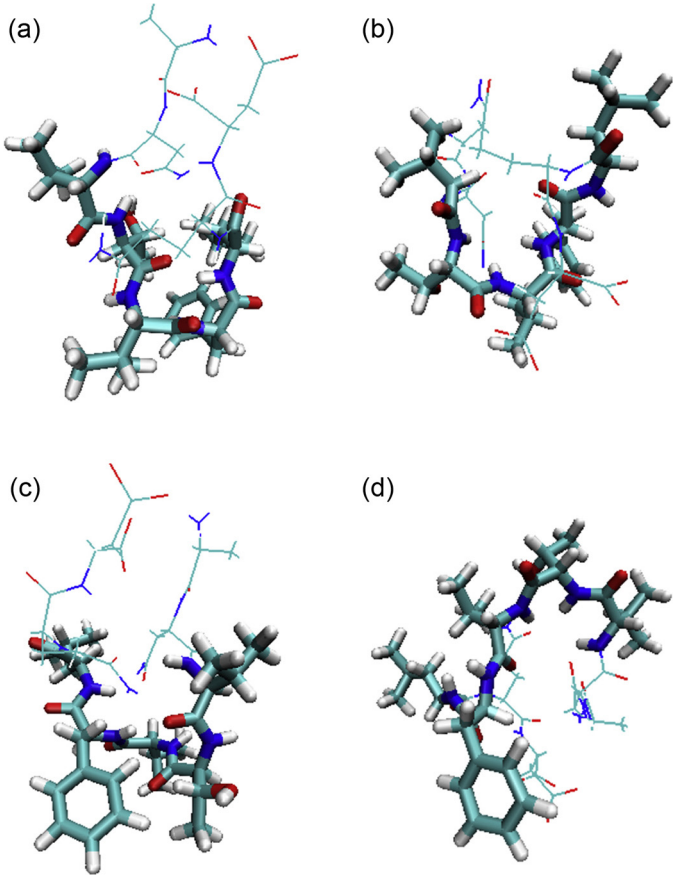

Fig. 3. Snapshots of four different folded structures: (a) CBS-9a, (b) CBS-9b, (c) CBS-9c and (d) CBS-9d. The thick portions represent the central -VTVFL- sequence of CBS-9, while the thin portions represent the two residues of each end of the peptide (AN- and -QD).

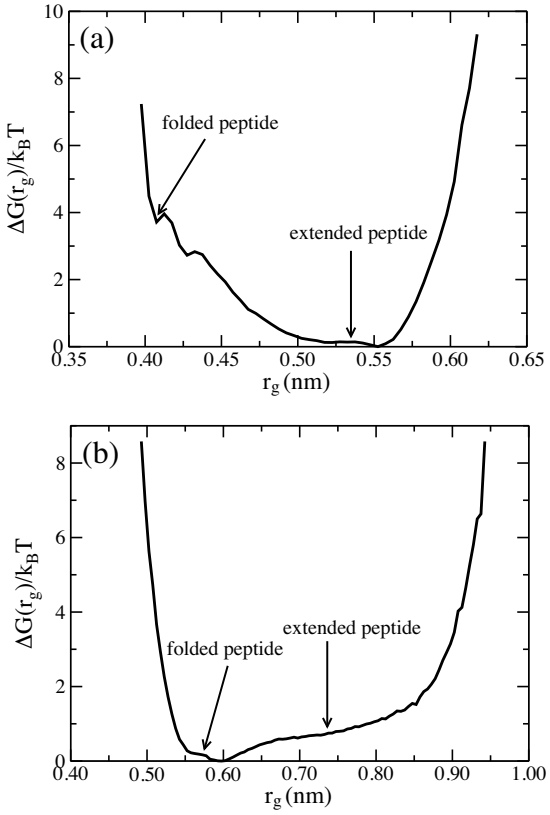

Fig. 4. The relative free energy as a function of r_g at $T = 37\,^{\circ}C$ for (a) CBS-5, for which extended states are preferred, and (b) CBS-9, for which folded and extended states are approximately equally likely.

one, but L is more frequent) and 3 (7.5%) of them bind to a narrow hydrophobic pocket at another location on CaM, shown in Fig. 5(c). The minority alternate binding arrangements tend to occur with zscores below 1120, which we shall see is a value associated with low binding energies.

The docking calculations on CBS-5x serve as a control for the idea that starting with folded configurations is better. CBS-5x produces the lowest top docking zscore out of all the peptide conformers. The top docking zscore reported by ZDOCK for each conformer is given in Table 2.

Table 2
Best zscore and calculated binding energy for each peptide conformer. Entries are listed by descending zscores for a given peptide length. Uncertainty in the binding energy is 0.7 kcal/mol at the 95% level.

Peptide	zscore	ΔG_b (kcal/mol)
CBS-5c	1236.4	12.2
CBS-5b	1159.6	8.6
CBS-5a	1129.6	9.7
CBS-5x	1119.7	5.4
CBS-9a	1219.5	7.9
CBS-9c	1201.6	7.3
CBS-9d	1191.8	5.9
CBS-9b	1123.1	5.6

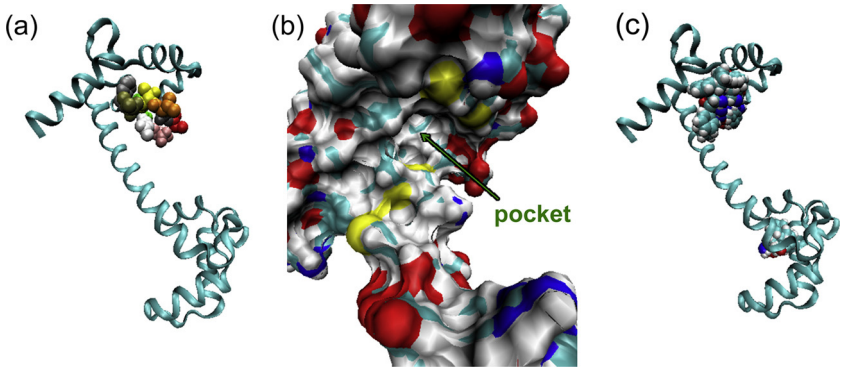

Fig. 5. (a) Representative docked configuration obtained from ZDOCK (specifically, highest scoring CBS-9a). This docking site is favoured across all CBS-5 and CBS-9 conformers. The colour code used for the peptide is as follows, A: red, N: dark grey, V: orange, T: yellow, V: gold, F: grey, L: green, Q: white and D: pink. CaM appears in a ribbon representation. (b) A close-up view of the docking area on CaM that features a hydrophobic pocket, with standard atomic colouring. (c) F residues for the top five docked configurations for each of the eight conformers studied. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

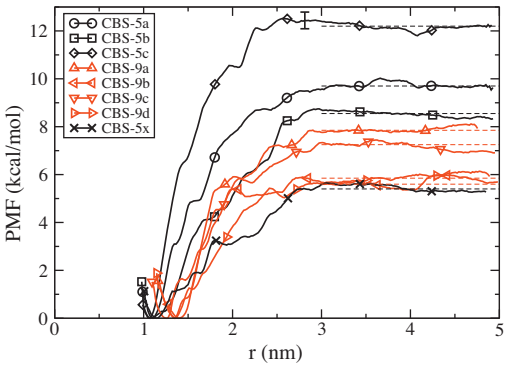

Fig. 6. The potential of mean force as a function of distance from the binding site calculated for the four CBS-5 (black) [three folded and one extended] and the four CBS-9 (red) peptides. An error bar representing a 95% confidence interval of the mean value is drawn for the curve for CBS-5c. The uncertainty is similar for all conformers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

3.3. Binding energy

We now proceed to use the top docked configuration for each of the eight candidate structures (which all fall in the same docking location on CaM) as starting points for US MD simulations that determine the binding free energy of the peptide to CaM. While the ZDOCK program evaluates binding of rigid peptide and protein structures, MD simulations allow for peptide and protein flexibility and thermal motion in order to determine the free energy. True equilibration of such simulations is computationally extremely demanding. Our approach on estimating the binding energy therefore rests on using the different independent starting configurations in order to help sample the phase space available to the peptide and protein while docking.

Fig. 6 shows the potential of mean force (PMF) as a function of distance between the centre of mass of the peptide and an atom on residue number 12 (phenylalanine) of CaM located slightly behind the hydrophobic pocket. Curves for each of the CBS-9 and CBS-5 conformers show a plateau, indicating a regime where the peptide is free of CaM. The height of the plateau, estimates of which are indicated by dashed lines, gives the binding free energy ΔG_b. Despite the good overlap between adjacent distance histograms (not shown) there appears to be still some noise in determining the PMF.

Fig. 7 shows snapshots of CBS-9a from simulation replicas with constraint distances of 1.3 nm (at the optimal peptide-binding site distance), 1.9 nm, 2.5 nm and 3.1 nm (where the peptide is nearly free of interaction with CaM). We emphasize that the constraints employed in the free energy calculation relate only to distance, and thus the peptide is free to sample positions within a spherical shell centred on the binding site, i.e., it is not constrained to follow a particular path out of the binding site but rather adopts paths of low free energy. The snapshots show that the peptide maintains contact with CaM when it is near the binding site. This "crawling" of the peptide along the CaM to get to the binding site contributes to the bumpiness of the PMF curves in Fig. 6. The many different microscopic pathways along which the peptide can approach the binding site, and the computational difficulty in sampling them, contribute to the noise in the PMF curves and to the uncertainty in binding energy estimates.

Note that the optimal distance between peptide and binding site is larger for CBS-9 simply because the centre of mass is used to calculate distance and CBS-9 is longer than CBS-5, shifting the centre of mass away from the binding site.

Table 3
Frequencies of contact of CaM residues with the binding site are obtained from a listing of the ten closest residues to each of V, T, V, F and L for each of the seven conformers of CBS-5 and CBS-9.

Residue ID	Frequency
47M	25
71K	21
67M	21
68M	20
32M	20
28L	20
15F	20
64F	19
51V	16
50E	13
48I	10
59I	8
23I	3

Estimates of ΔG_b for the CBS-9a-d peptides are 7.9, 5.6, 7.3 and 5.9 kcal/mol, respectively, for an average of 6.7 kcal/mol. The estimates for CBS-5a-c peptides are consistently higher at 9.7, 8.6, and 12.2 kcal/mol, respectively, with an average of 10.2 kcal/mol. For the extended version, CBS-5x, the binding energy is 5.4 kcal/mol, by far the lowest of the CBS-5 conformers. WHAM analysis gives an uncertainty estimate for the binding energy of 0.7 kcal/mol at the 95% level (twice the standard deviation in the mean). The binding energies for the peptides are listed in Table 2.

3.4. CaM binding site residues

In an effort to identify the important residues in CaM, we calculate the average distance between the centres of mass of each of the amino acids (aa's) V_1, T, V_2, F and L in the peptides and all of the CaM residues. We do this for each of the seven binding energy simulations we have of the folded peptides (CBS-5a-c and CBS-9a-d) for which the peptide is restrained to be nearest the optimal binding distance (being the minimum of the PMF curve), e.g., the simulation with $r_0 = 1.3$ nm for CBS-9a and $r_0 = 1.15$ nm for all CBS-5 peptides. We use the ensemble of configurations from these optimal-distance replicas to calculate average distances between residues in the peptides and those in CaM. We characterize the CaM residues surrounding the seven docked candidates in two ways.

(1) For each of the 5 aa's in each of the 7 peptide candidates, we find the 10 closest CaM residues. This produces a list of 350 CaM residue identities, many of which are repeated. There are 37 distinct residues. We rank these according to the number of times they appear in the list of 350. We visually examined each of these 37 to determine which of them are in "contact" with the peptide. Some residues that have a fairly high frequency are cut from the list because they are second nearest neighbours to the peptide, i.e., they are next to residues that are in contact with the peptide. There are 13 residues in contact, and they are listed in Table 3.

If the list of 350 residues is broken down for each of V_1, T, V_2, F and L, we find that the list for F has only 16 distinct residues out of the 70 in total, while the other aa's have around 30 distinct residues. This means that the environment of F is much less variable. This makes sense as F sits most deeply in the hydrophobic pocket of the binding site.

(2) Given the importance of F, which anchors most deeply into the docking site, we find the 13 calmodulin residues closest on average to F (averaged over the seven folded conformers). They are listed in Table 4.

The two lists, the "frequency list" in Table 3 and the "F distance list" in Table 4 are the same, except that residue 12 appears in the F distance list and 32 appears instead in the frequency list. Thus, we judge the important 14 residues to be 12F, 15F, 23I, 28L, 32M,

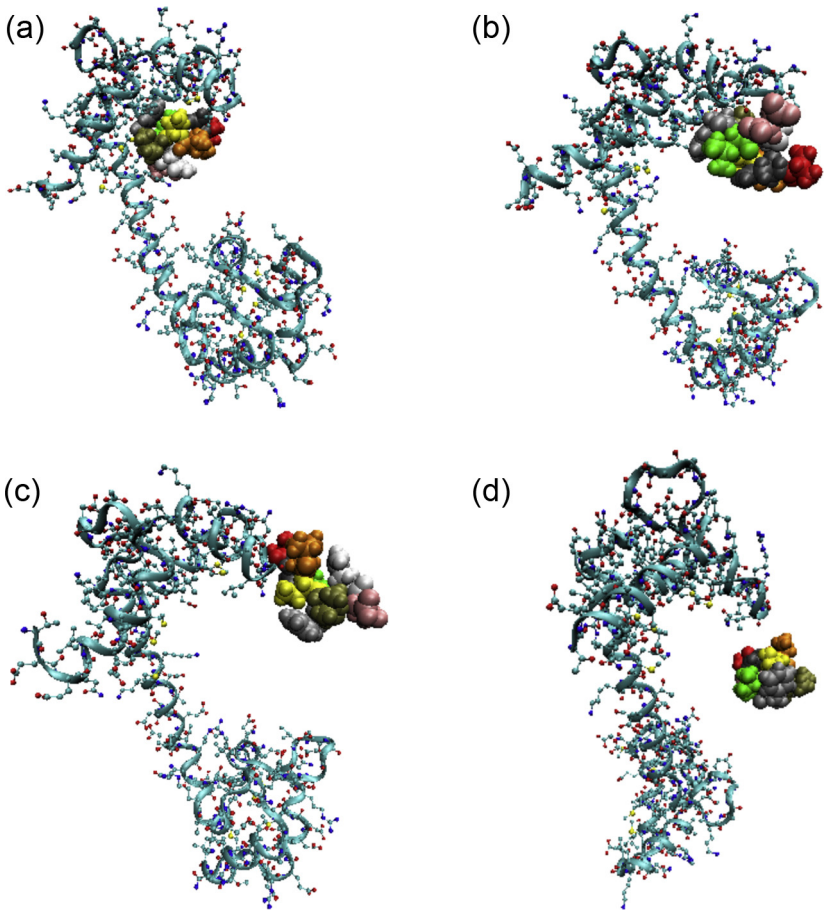

Fig. 7. Snapshots for the CBS-9a peptide pulled away from the calmodulin using the pull code provided by GROMACS from replicas with values of r_0 equalling (a) 1.3 nm (b) 1.9 nm (c) 2.5 nm and (d) 3.1 nm. Peptide colouring is as in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Table 4
Distances of the thirteen closest CaM residues to the phenylalanine, averaged over the seven peptide conformers.

Residue ID	Distance to F (nm)
47M	0.663
64F	0.677
67M	0.710
51V	0.728
15F	0.738
28L	0.740
68M	0.773
48I	0.796
59I	0.797
23I	0.875
50E	0.891
71K	0.893
12F	0.911

47M, 48I, 50E, 51V, 59I, 64F, 67M, 68M, 71K, which in context of the 143-residue calmodulin sequence reads as (identified residues in red),

TEEQIAEFKEAF SLFDKDGDGTIT TKELGTVMRSLG
QNPTEAELQDMI NEVDADGNGTID FPEFLTMMARKM
KDTDSEEEIREA FRVFDKDGNGYI SAAELRHVMTNL
GEKLTDEEVDEM IREANIDGDGQV NYEEFVQMMTA.

The deepest part of the binding site, hydrophobic pocket, may be considered to be, as determined by close visual inspection, formed by the side chains of residues 15F, 23I, 28L, 48I, 51V, 59I and 64F. Fig. 8(a) shows a snapshot of CaM highlighting the 14 residues that form the docking site with and without the peptide in place. The docking site is lined with hydrophobic groups, consistent with the hydrophobic aa residues exposed when the peptide folds. Four methionines are immediately apparent. Some of the CaM residues that are not highlighted but appear to be close to the peptide can be characterized as flexible or hinged appendages. The position of these "claws" varies in time and their role in docking is unclear. The "peripheral" residues of the docking site also show a degree of flexibility and are therefore somewhat accommodating to different peptide conformations [40].

Knowing which residues form the binding pocket provides an avenue for rational drug design, particularly through mutagenesis studies [41]. Identifying the binding pocket through a simple distance criterion, such as finding residues that have at least one non-hydrogen atom within 0.5 nm of a non-hydrogen atom of the ligand, is appealing and straightforward. This method was originally used to define the binding pocket of ATP in the Cdk5-Nck5a* complex [42] and has been used in multiple drug design studies [4,5,43]. However, we hope that our approach is useful when there is a variety of different peptide structures that the binding site can accommodate, as in the present case.

3.5. Bound peptide structure

We now examine in more detail the configurations that the peptide takes during the US simulations at the optimal binding distance in order to gain insight into the factors that may influence the binding energy. While we see significant structural change of

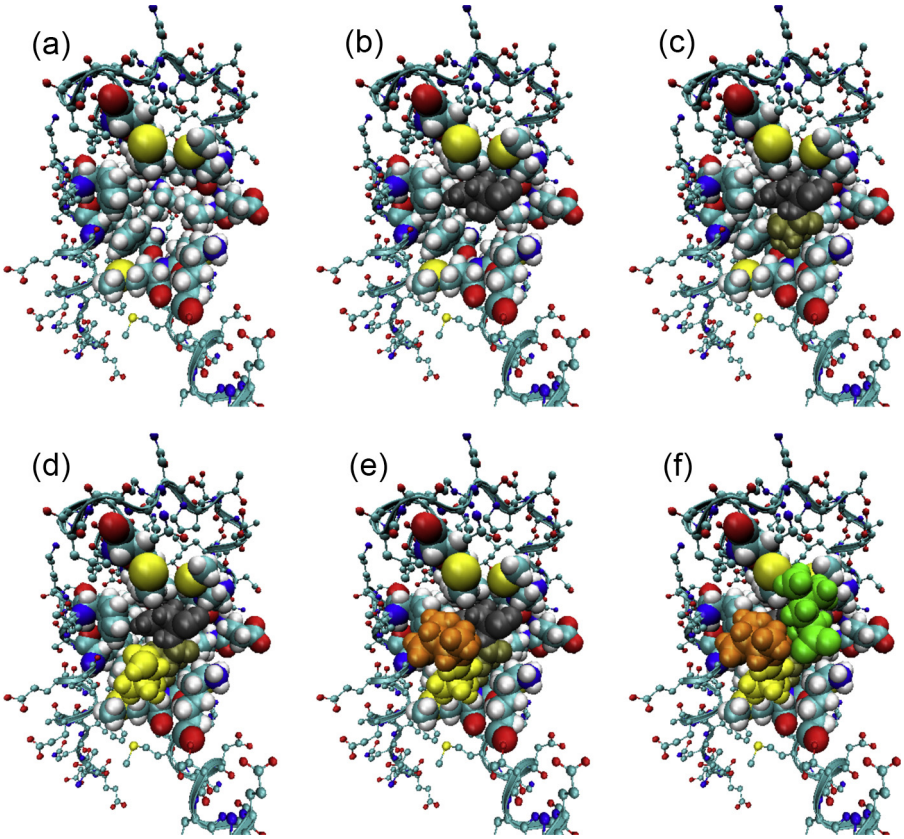

Fig. 8. Snapshot of CaM with bound CBS-5c from US simulations. The 14 CaM residues considered important for binding are represented with spheres of standard colouring: 12F, 15F, 23I, 28L, 32M, 47M, 48I, 50E, 51V, 59I, 64F, 67M, 68M, 71K. Panels show progressively more of CBS-5c, showing the most deeply embedded aa residues first: (a) just the binding site without a CBS-5 peptide in place, (b) F [grey], (c) V_2 [gold], (d) T [yellow], (e) V_1 [orange] and (f) L [green].

the peptide in Fig. 1 occurring well within 50ns, the much more confined environment around the bound peptide means that we are determining the binding energy for more or less a particular conformation of the peptide during the US simulations.

In Fig. 8, we show the locations of the aa residues for a representative snapshot of CBS-5c, the conformer with the largest binding energy, bound to CaM from the US simulation at the optimal distance. Fig. 8(b) shows that the side chain of F is most deeply embedded into the hydrophobic pocket of the binding site and Fig. 8(c) shows that the side chain of the second valine in the peptide sequence, V_2, accompanies F into the pocket. This adjacency of F and V_2 and their position within the pocket is preserved from the initially chosen conformation [as in Fig. 2(c)]. Fig. 8(d)–(f) shows the remainder of the aa residues, with L in particular resting snugly in the hydrophobic periphery of the binding site along with V_1, the first valine.

Consideration of the other six folded conformers and the extended version of CBS-5, which we detail below, indicates the importance of the V_2–F pairing in producing a high binding energy.

In contrast to CBS-5c, both initial configurations of CBS-5a and CBS-5b, which are nearly identical to each other [see Fig. 2(a) and (b)], have a pairing of T and F, rather than V_2 and F, and yield (similar) lower binding energies. The T–F pairing persists during US simulations for CBS-5a. The case of CBS-5b is interesting in that during the US simulation, there is a conformational change resulting in a pairing of V_2 and F, as in CBS-5c. The very fact that this conformational change occurs is consistent with the V_2–F pairing being thermodynamically more stable than the T–F pairing. We note that this change does not result in a larger binding energy for CBS-5b, as the side chains of V_2 and F do not achieve the same level of proximity as for CBS-5c and hence do not fit as well into the pocket. For both CBS-5a and CBS-5b, as for CBS-5c, V_1 and L remain in close proximity to each other and nestle snugly in the hydrophobic periphery of the binding site.

For the extended version CBS-5x, only F acts as an anchor. Over the 50 ns of the US simulations, the peptide does optimize its configuration near the binding site to a degree, but still remains extended, with V_1 and L far apart. The time needed to relax fully is significantly longer, and therefore CBS-5x still produces the lowest binding free energy out of all CBS-5 conformers. This validates using folded conformations as starting points for binding calculations.

CBS-9a, CBS-9b and CBS-9d are similar to one another and to CBS-5x in that only F anchors the peptide to CaM, with both V_2 and T side chains situated away from the pocket portion of the binding site. For these three conformers, L remains close to the pocket, similarly to CBS-5a-c. In addition, CBS-9a, which has the highest binding energy of the CBS-9 conformers, maintains V_1 and L close to each other and to the pocket. For CBS-9c, all three side chains of T, V_2 and F are situated in the pocket, but not as snugly as when only two side chains enter the pocket. As well, L for CBS-9c is further from the pocket. The lack of a proper co-anchor to F in all versions of CBS-9 is consistent with their lower binding energies.

4. Discussion and conclusions

In this study, we take the approach of finding several folded conformational variants of two sub-peptides of different length (CBS-5 and CBS-9) of a potentially therapeutic peptide shown to act on calmodulin. Treating both CaM and the conformers initially as rigid objects, we find the optimal docked configurations, with all conformers preferring the same binding site on CaM. We then use these docked configurations as starting points for US

binding energy calculations. These US simulations are long enough to provide reasonable estimates of the binding energy, but not long enough to allow for significant structural relaxation of the peptide while it is near the binding site. This allows us to correlate differences in binding energy to structural differences in how the peptide binds to CaM.

The binding site itself has a deep hydrophobic pocket as well as hydrophobic periphery. The pocket provides an anchoring site, and while phenyl groups are known to act as anchors for disordered peptides [30], the conformation of neighbouring aa residues also plays a role [29]. In our case, the strongest binding occurs when the side chains of F and V_2 act together as the anchor. Weaker binding occurs when F and T act together, while F acting alone produces the weakest binding.

A secondary factor is the folding of the peptide that results in V_1 and L being next to each other, allowing them to fit along the periphery of the biding site near the pocket. This also seems to enhance binding, as it is a common feature across folded 5 aa conformers CBS-5a, CBS-5b, CBS-5c and CBS-9a, the best of the 9 aa conformers in terms of binding free energy.

Our findings are consistent with the claims of the patent that the important part of the peptide sequence is V X_1 X_2 F L with X_1 hydrophobic or neutral (e.g., V, I, L, etc., but typically T, A, P) and similarly for X_2, typically V or T.

CBS-5 tends not to fold in water only, and compact, internally hydrogen-bonded structures are rare. Of the three most compact structures we found in simulations with water only, two were nearly identical, indicating a small number of available folded conformations. Despite this propensity to remain extended, CBS-5 generally has a higher binding energy compared to CBS-9, which tends to fold to many available structures. That the smaller peptide produces better binding is encouraging for future development of small non-peptide analogues as orally administered drugs.

Acknowledgments

We thank Drs. Valerie K. Booth and Marek W. Bromberek of Memorial University of Newfoundland for helpful discussions. We would like to acknowledge NSERC (Canada) for funding, in particular the Interact and Engage programs. Computational facilities are provided by ACENET, the regional high performance computing consortium for universities in Atlantic Canada. ACENET is funded by the Canada Foundation for Innovation (CFI), the Atlantic Canada Opportunities Agency (ACOA), and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick.

References

[1] M. Husain, J. Choi, Calmodulin-binding peptides that reduce cell proliferation in cancer and smooth muscle proliferation diseases, International Application Number PCT/CA2008/001474, International Publication Number WO 2009/023959 A1, February 26, 2009, http://www.google.com/patents/WO2009023959A1

[2] M.J. Berardi, W.M. Shih, S.C. Harrison, J.J. Chou, Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching, Nature 476 (2011) 109–113, http://dx.doi.org/10.1038/nature10257.

[3] B. OuYang, S. Xie, M.J. Berardi, X. Zhao, J. Dev, W. Yu, B. Sun, J.J. Chou, Unusual architecture of the p7 channel from hepatitis c virus, Nature 498 (2013) 521–525, http://dx.doi.org/10.1038/nature12283.

[4] J.-F. Wang, D.-Q. Wei, L. Li, S.Y. Zheng, Y.-X. Li, K.-C. Chou, 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design, Biochem. Biophys. Res. Commun. 355 (2007) 513–519, http://dx.doi.org/10.1016/j.bbrc.2007.01.185 (Corrigendum: Biochem. Biophys. Res. Commun. 357 (2007) 330).

[5] Y. Ma, S.-Q. Wang, W.-R. Xu, R.-L. Wang, K.-C. Chou, Design novel dual agonists for treating type-2 diabetes by targeting peroxisome proliferator-activated receptors with core hopping approach, PLoS ONE 7 (2012) e38546, http://dx.doi.org/10.1371/journal.pone.0038546.

[6] K.-C. Chou, Impacts of bioinformatics to medicinal chemistry, Med. Chem. 11 (2015) 218–234, http://dx.doi.org/10.2174/1573406411666141229162834.

[7] K.-C. Chou, Low-frequency collective motion in biomacromolecules and its biological functions, Biophys. Chem. 30 (1988) 3–48, http://dx.doi.org/10.1016/0301-4622(88)85002-6.

[8] K.-C. Chou, Low-frequency resonance and cooperativity of hemoglobin, Trends Biochem. Sci. 14 (1989) 212–213, http://dx.doi.org/10.1016/0968-0004(89)90026-1.

[9] J.-F. Wang, K. Gong, D.-Q. Wei, Y.-X. Li, K.-C. Chou, Molecular dynamics studies on the interactions of PTP1B with inhibitors: from the first phosphate-binding site to the second one, Protein Eng. Des. Sel. 22 (2009) 349–355, http://dx.doi.org/10.1093/protein/gzp012.

[10] K.-C. Chou, B. Mao, Collective motion in DNA and its role in drug intercalation, Biopolymers 27 (1988) 1795–1815, http://dx.doi.org/10.1002/bip.360271109.

[11] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447, http://dx.doi.org/10.1021/ct700301q.

[12] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible and free, J. Comput. Chem. 26 (2005) 1701–1718, http://dx.doi.org/10.1002/jcc.20291.

[13] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317, http://dx.doi.org/10.1007/s008940100045.

[14] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56, http://dx.doi.org/10.1016/0010-4655(95)00042-E.

[15] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25 (2004) 1656–1676, http://dx.doi.org/10.1002/jcc.20090.

[16] Y.S. Babu, C.E. Bugg, W.J. Cook, Structure of calmodulin refined at 2.2 A resolution, J. Mol. Biol. 204 (1988) 191–204, http://dx.doi.org/10.1016/0022-2836(88)90608-0.

[17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acid Res. 28 (2000) 235–242, http://dx.doi.org/10.1093/nar/28.1.235.

[18] PDB ID: 3clndoi:10.2210/pdb3cln/pdb. http://www.rcsb.org

[19] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725, http://dx.doi.org/10.1002/prot.21123.

[20] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew, J. Chem. Phys. 120 (2004) 9665–9678, http://dx.doi.org/10.1063/1.1683075.

[21] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer. An environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723, http://dx.doi.org/10.1002/elps.1150181505.

[22] W. Humphrey, A. Dalke, K. Schulten, VMD. Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, http://dx.doi.org/10.1016/0263-7855(96)00018-5.

[23] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255–268, http://dx.doi.org/10.1080/00268978400101201.

[24] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A 31 (1985) 1695–1697, http://dx.doi.org/10.1103/PhysRevA.31.1695.

[25] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190, http://dx.doi.org/10.1063/1.328693.

[26] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, Mol. Phys. 50 (1983) 1055–1076, http://dx.doi.org/10.1080/00268978300102851.

[27] B.G. Pierce, Y. Hourai, Z. Weng, Accelerating protein docking in ZDOCK using an advanced 3D convolution library, PLoS ONE 6 (2011) e24657, http://dx.doi.org/10.1371/journal.pone.0024657 zdock.umassmed.edu/software

[28] J. Mintseris, B. Pierce, K. Wiehe, R. Anderson, R. Chen, Z. Weng, Integrating statistical pair potentials into protein complex prediction, Proteins 69 (2007) 511–520, http://dx.doi.org/10.1002/prot.21502.

[29] Y. Huang, Z. Liu, Anchoring intrinsically disordered proteins to multiple targets: lessons from N-terminus of the p53 protein, Int. J. Mol. Sci. 12 (2011) 1410–1430, http://dx.doi.org/10.3390/ijms12021410.

[30] I. Staneva, Y. Huang, Z. Liu, S. Wallin, Binding of two intrinsically disordered peptides to a multi-specific protein: a combined Monte Carlo and molecular dynamics study, PLoS Comput. Biol. 8 (2012) e1002682, http://dx.doi.org/10.1371/journal.pcbi.1002682.

[31] J.A. Lemkul, D.R. Bevan, Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics, J. Phys. Chem. B 114 (2009) 1652–1660, http://dx.doi.org/10.1021/jp9110794.

[32] G.N. Patey, J.P. Valleau, The free energy of spheres with dipoles: Monte Carlo with multistage sampling, Chem. Phys. Lett. 21 (1973) 297–300, http://dx.doi.org/10.1016/0009-2614(73)80139-3.

[33] G.M. Torrie, J.P. Valleau, Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard–Jones fluid, Chem. Phys. Lett. 28 (1974) 578–581, http://dx.doi.org/10.1016/0009-2614(74)80109-0.

[34] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, J. Comput. Phys. 23 (1977) 187–199, http://dx.doi.org/10.1016/0021-9991(77)90121-8.

[35] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021, http://dx.doi.org/10.1002/jcc.540130812.
[36] J.S. Hub, B.L. de Groot, D. van der Spoel, g_wham, a free weighted histogram analysis implementation including robust error and autocorrelation estimates, J. Chem. Theory Comput. 6 (2010) 3713–3720, http://dx.doi.org/10.1021/ct100494z.
[37] L. Li, D.-Q. Wei, J.-F. Wang, K.-C. Chou, Computational studies of the binding mechanism of calmodulin with chrysin, Biochem. Biophys. Res. Commun. 358 (2007) 1102–1107, http://dx.doi.org/10.1016/j.bbrc.2007.05.053.
[38] C. Yang, K. Kuczera, Molecular dynamics simulations of a calmodulin–peptide complex in solution, J. Biomol. Struct. Dyn. 20 (2002) 179–197, http://dx.doi.org/10.1080/07391102.2002.10506834.
[39] K. Kuczera, P. Kursula, Interactions of calmodulin with death-associated protein kinase peptides: experimental and modeling studies, J. Biomol. Struct. Dyn. 30 (2012) 45–61, http://dx.doi.org/10.1080/07391102.2012.674221.
[40] J.-H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, Computational drug design accommodating receptor flexibility: the relaxed complex scheme, J. Am. Chem. Soc. 124 (2002) 5632–5633, http://dx.doi.org/10.1021/ja0260162.
[41] K.-C. Chou, Structural bioinformatics and its impact to biomedical science, Curr. Med. Chem. 11 (2004) 2105–2134, http://dx.doi.org/10.2174/0929867043364667.
[42] K.-C. Chou, K.D. Watenpaugh, R.L. Heinrikson, A model of the complex between cyclin-dependent kinase 5 and the activation domain of neuronal Cdk5 activator, Biochem. Biophys. Res. Commun. 259 (1999) 420–428, http://dx.doi.org/10.1006/bbrc.1999.0792.
[43] K.-C. Chou, D.-Q. Wei, W.-Z. Zhong, Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS, Biochem. Biophys. Res. Commun. 308 (2003) 148–151, http://dx.doi.org/10.1016/S0006-291X(03)01342-1 (Erratum: Biochem. Biophys. Res. Commun. 310 (2003) 675).