BIOINFORMATICS ORIGINAL PAPER Vol. 27 no. 13 2011, pages 1814–1821
doi:10.1093/bioinformatics/btr294

Structural bioinformatics

Advance Access publication May 17, 2011

Prediction of peptides binding to the PKA RIIα subunit using a hierarchical strategy

Tingjun Hou[1,*], Youyong Li[1] and Wei Wang[2,*]

[1]Institute of Functional Nano & Soft Materials (FUNSOM) and Jiangsu Key Laboratory for Carbon-Based Functional Materials & Devices, Soochow University, Suzhou, Jiangsu 215123, P. R. China and [2]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093-0359, USA

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Favorable interaction between the regulatory subunit of the cAMP-dependent protein kinase (PKA) and a peptide in A-kinase anchoring proteins (AKAPs) is critical for translocating PKA to the subcellular sites where the enzyme phosphorylates its substrates. It is very hard to identify AKAPs peptides binding to PKA due to the high sequence diversity of AKAPs.

Results: We propose a hierarchical and efficient approach, which combines molecular dynamics (MD) simulations, free energy calculations, virtual mutagenesis (VM) and bioinformatics analyses, to predict peptides binding to the PKA RIIα regulatory subunit in the human proteome systematically. Our approach successfully retrieved 15 out of 18 documented RIIα-binding peptides. Literature curation supported that many newly predicted peptides might be true AKAPs. Here, we present the first systematic search for AKAP peptides in the human proteome, which is useful to further experimental identification of AKAPs and functional analysis of their biological roles.

Contact: tingjunhou@hotmail.com; tjhou@suda.edu.cn; wei-wang@ucsd.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on January 21, 2011; revised on May 3, 2011; accepted on May 4, 2011

1 INTRODUCTION

It is critical to understand how the functions of kinases are regulated. A classic kinase is cAMP-dependent protein kinase, also known as protein kinase A (PKA) (Kim *et al*., 2005; Langeberg and Scott, 2005). PKA is a heterotetramer composed of two catalytic (C) subunits and two regulatory (R) subunits. The regulatory subunit (R) has four isoforms, RIα, RIβ, RIIα and RIIβ. Each R subunit contains an N-terminal dimerization/docking (D/D) domain followed by an inhibitor site and two tandem cAMP binding domains. The D/D domain is a four anti-parallel helix bundle, which provides a stable and hydrophobic surface for peptides in A-kinase anchoring proteins (AKAPs). The protein–protein interaction between the PKA R subunit dimer and AKAP makes the enzyme locate at the subcellular sites where it phosphorylates the substrates to initiate downstream signal transduction (Carnegie and Scott, 2003; Langeberg and Scott, 2005). It is no doubt that the identification of AKAPs is critical for understanding the functional roles of PKA.

AKAPs share similar functional properties but have quite diverse sequences and structures (Appert-Collin *et al*., 2006), which makes it extremely difficult to identify AKAP proteins solely based on sequence or structural similarity (Alto and Scott, 2004). Consequently, only a limited number of AKAPs have been identified so far in human (Appert-Collin *et al*., 2006).

Numerous computational approaches have been developed to predict the binding peptides of modular domains, such as SH2, SH3, WW and PDZ (Brannetti *et al*., 2000; Ferraro *et al*., 2006; Fong *et al*., 2004; Hou *et al*., 2006a, b, 2008, 2009; McLaughlin *et al*., 2006; Obenauer *et al*., 2003; Schleinkofer *et al*., 2004). The success of these approaches on various systems demonstrated the efficiency of the computational methods to predict the transient and weak binding between protein domains and peptides. Since there are not many known AKAP peptides and no sequence motif for AKAP peptides has been detected, it is not straightforward to apply these methods to identify PKA–AKAP interactions. Here, we developed a hierarchical approach by integrating information of binding free energy calculation, secondary structure prediction, transmembrane (TM) prediction and sequence conservation analyses of candidate peptides to predict putative AKAPs in the human genome. Since only crystal structure between AKAP peptide and PKA RIIα is available, here we focused on predicting AKAPs specific for RIIα. Our approach is readily applicable to identify the binding partners of other regulatory subunit isoforms of PKA as well.

Our approach includes the following steps: (i) The SWISSPROT database (Apweiler *et al*., 2004) was screened to find potential AKAP peptides using a position specific scoring matrix (PSSM) derived from the Virtual Mutagenesis (VM; Hou *et al*., 2006a, b) results and prior knowledge; (ii) Secondary structure for each candidate peptide was predicted and the peptides without helices were removed (17 of the 18 known AKAP peptides were predicted to form α helices); (iii) TM domain for each candidate peptide was predicted and the peptides located within TM regions were removed (all known AKAP peptides were predicted outside TM regions); and (iv) Conservation analysis across seven vertebrate species was conducted on each remaining peptide and the nonconserved ones were removed (all known AKAP peptides are conserved). The top 500 putative AKAP peptides predicted from the human proteome included the 14 known ones. Literature curation also suggested that our predictions are able to identify new AKAP peptides. We believe that this systematic prediction of AKAP peptides from the

*To whom correspondence should be addressed.

entire human proteome provides the critical information for further experimental identification of AKAPs and investigation of their biological functions.

2 METHODS

2.1 Molecular dynamics simulations

The crystal structure of the AKAP10 (also known as D-AKAP2) peptide complexed with the RIIα D/D domain (PDB entry: 2hwn) was used as the template (Kinderman *et al*., 2006). The 1.6 Å template of the D/D domain bound to a 22-residue AKAP peptide from AKAP10 has two dimers–AKAP complexes in the asymmetric unit. Each monomer of the D/D domain is 39-residue long. The binding peptide in the complex is 20-residue long, and the peptide positions from the N- to C-terminus were referred as P2–P21 (the residues at P1 and P22 are missing in the crystal structure). The first dimers–AKAP complex in the crystal structure was used to build the PSSM. The two dimer–AKAP complexes align with an rmsd of 0.21 Å for the Cα atoms or an rmsd of 0.24 Å for the main-chain atoms. It is obvious that these two dimer–AKAP complexes are quite similar. The complex was solvated in a rectangular box of TIP3P water with a 9 Å buffer between the solute and the box edge. The Particle Mesh Ewald (PME) method was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images (Darden *et al*., 1993). AMBER03 force field was used in the simulation (Duan *et al*., 2003). Before molecular dynamics (MD) simulations, the complex was gradually relaxed by 5000 cycles of minimization procedure (500 cycles of steepest descent and 4500 cycles of conjugate gradient minimization). MD simulations were performed in the NPT ensemble at a constant temperature of 300K and a constant pressure of 1 atm. The SHAKE procedure was employed to constrain all hydrogen atoms (Ryckaert *et al*., 1977), and the time step was set to 2.0 fs. The system was first heated in the NVT ensemble from 10K to 300K over 20 ps. In the following data collection period, snapshots were saved every 4 ps, yielding a total of 250 snapshots, and the last 200 snapshots were used for free energy calculation. All MD simulations were performed using the *pmemd* program of AMBER9 (Case *et al*., 2005).

2.2 The MM/PBSA calculations and the MM/GBSA free energy decomposition analysis

The binding free energy (ΔG_{bind}) was calculated using the MM/PBSA method (Kollman *et al*., 2000; Wang *et al*., 2001, 2006):

$$\begin{aligned}\Delta G_{\text{bind}} &= G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \\ &= \Delta E_{MM} + \Delta G_{PB} + \Delta G_{SA} - T\Delta S\end{aligned} \quad (1)$$

where ΔE_{MM} represents the change of molecular mechanics potential energy upon ligand binding; ΔG_{PB} and ΔG_{SA}represent the electrostatic component and the nonpolar component of solvation free energy, respectively; $-T\Delta S$ is the conformational entropy change.

ΔG_{PB} was calculated using the *pbsa* program in AMBER9. The grid size used to solve the Poisson–Boltzmann equation was 0.5 Å, and the values of interior dielectric constant and exterior dielectric constant were set to 1 and 80, respectively. ΔG_{SA} was estimated based on the surface area: $0.0072 \times \Delta$SASA. The peptide–RIIα interaction energies were calculated using 200 snapshots taken from 0.2 ns to 1.0 ns MD simulation trajectories of the complex. The normal mode analysis was performed to estimate the vibrational component of the entropy using the *nmode* program in AMBER9.0 (Case *et al*., 2005). Since the normal mode analysis is computationally expensive, only 50 snapshots were used to estimate –TΔS.

The interaction between each residue in the RIIα D/D domain and each peptide residue was computed using the MM/GBSA decomposition protocol implemented in the *mm_pbsa* module in AMBER9. The binding interaction of each residue–residue pair includes three terms: van der Waals contribution (ΔG_{vdw}), electrostatic contribution (ΔG_{ele}) and solvation contribution ($\Delta G_{\text{solvation}}$). The polar contribution (ΔG_{GB}) of $\Delta G_{\text{solvation}}$ was computed using generalized Born model based on the parameters developed by Onufriev *et al.* (2000) All energy components were calculated using 50 snapshots taken from the 0.2 ns to 1.0 ns MD trajectory.

2.3 VM and PSSM

The energetic preferences of peptide residues at six positions, P5, P8, P9, P12, P13 and P16, were analyzed using the VM technique (Hou *et al*., 2006a) because peptide array experiments indicated the preference of these amino acids (Burns-Hamuro *et al*., 2003). The peptide in the crystal structure was used as the template and side chain mutation was performed by *scap* (Xiang and Honig, 2001). Because only small or hydrophobic residues were favored at these six positions as shown by the peptide array experiments (Burns-Hamuro *et al*., 2003), only ten amino acids, Ile, Gly, Leu, Met, Phe, Ser, Thr, Tyr, Val and Ala, were considered in VM. Minimization, MD simulations and MM/PBSA calculations were performed on the 10 mutated complexes for each position using the same setup described above. The difference between the binding free energies of the mutated peptide and the template peptide ΔΔG was calculated, which was used in a PSSM to represent the motif of the RIIα-binding peptides. For the amino acids not calculated by VM, an arbitrary value of 50 was used in the PSSM to penalize their presence. We considered the other peptide positions as noninformative and assigned a value of zero in the PSSM for all amino acids but proline. Because proline is highly unfavorable in helix, a larger penalty of 80 was assigned to proline at each position from P1 to P19. The PSSM was then used to score all 21-residue long human peptides in the SWISSPROT database (Apweiler *et al*., 2004). The score of each peptide was calculated as: $\sum_{i=1}^{21} M_{S,i}\delta(S,S_i)$, where $M_{S,i}$ is the score of the amino acid S at i-th position in the PSSM and S_i is the amino acid at the i-th position of the peptide (the PSSM is shown in Supplementary Table S1). For each mutated system, 1 ns MD simulations consumed about 96 (12×8) CPU hours at Xeon 5130 CPU. Therefore, the generation of the PSSM takes roughly 6.75 days with a 32-CPU Linux cluster that is becoming commonly available.

2.4 Secondary structure prediction

Because α helix is essential for a peptide to bind the RII subunit, secondary structure prediction was used as a filter to remove false positives. The secondary structures for the proteins that contain putative RII-binding peptides were predicted using three methods, Predator (Frishman and Argos, 1997), Jnet (Cuff and Barton, 2000) and Prof (Rost, 1996). Predator only needed the query sequence, but multiple sequence alignment for the query sequence was required for both Jnet and Prof. For the latter two methods, the top 20 homologs of the query sequence were found using PSI-BLAST (Altschul *et al*., 1997) from the BLAST nonredundant protein sequence database (http://www.ncbi.nlm.nih.gov/RefSeq/). Multiple sequence alignment was then generated using ClustalW1.7 (Thompson *et al*., 1994). The percentage of α-helix for the putative (S_{helix}) RIIα-binding peptide was calculated as: $S_{\text{helix}} = N/21$, where N is the number of residues predicted to be α-helical conformation. In order to have a more robust prediction for each peptide, we calculated a consensus score by averaging the best two S_{helix} scores obtained from the three methods. An arbitrary cutoff of 0.88 was chosen for the consensus score to determine whether a peptide can form an α-helix.

2.5 TM predictions

Because TM regions of a protein cannot interact with PKA, TM domain prediction was used as a filter to remove false positives. TM domains of a protein were predicted using the TMHMM program (Krogh *et al*., 2001). A candidate peptide will be removed if 10 residues of this peptide were predicted in the TM domain.

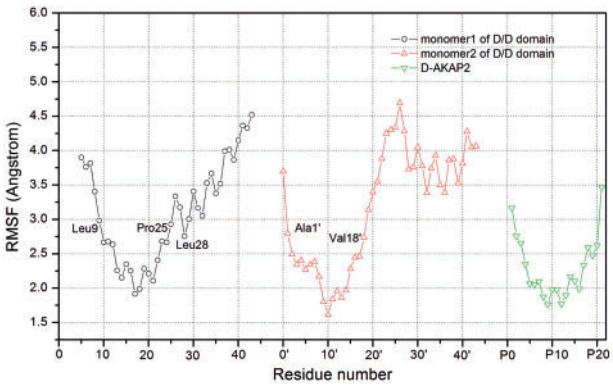

Fig. 1. RMSF of backbone atoms in the RIIα/AKAP10 complex. *X*-axis is the residue number (residues in monomer 1 are indicated by plain numbers and those in monomer 2 by primes).

2.6 Conservation analysis based on multiple sequence alignment

Because of the functional constraint, the peptides binding to RIIα should be more conserved than random peptides. Therefore, we conducted conservation analysis across seven vertebrate species for each putative binding peptide to remove false positives further. The seven species are *Homo sapiens* (human), *Mus musculus* (mouse), *Bos taurus* (cow), *Canis familiaris* (dog), *Rattus norvegicus* (rat), *Pan troglodytes* (chimpanzee) and *Macaca mulatta* (rhesus macaque). The protein sequences for the seven species were taken from the NCBI nonredundant protein sequence database. For each protein including the putative binding peptide, the best match in each species, was considered as a homolog if *E*-value $< 10^{-10}$; otherwise, no homolog was considered. The human protein sequence and its homologs were then aligned using ClustalW 1.7 (Thompson *et al*., 1994). Using this alignment, for the 21-residue long putative binding peptide, we calculated a pair-wise similarity score: $\sum_{i=1}^{21} S_{A_iB_i} / \sum_{i=1}^{21} S_{A_iA_i}$, where $S_{A_iB_i}$ is the amino acid similarity score in the PAM500 mutation matrix (Altschul, 1991) between residue A at the i-th position in the human peptide and residue B_i in the other species. The penalty for gap was set to −10. If no homolog was found, we considered that the conservation analysis was not informative and set the sequence similarity to 1.0. The average similarity score across the seven species was used to evaluate the degree of conservation of a putative RIIα-binding peptide.

3 RESULTS

3.1 The interactions between AKAP10 and the RIIα D/D domain

Our 1 ns MD simulation results provided the dynamic details of the molecular interactions between the PKA regulatory subunit and its binding peptides. We first confirmed that the MD simulations were stable and equilibrated by monitoring the root mean square deviation (RMSD) of all backbone atoms relative to the starting structure (Supplementary Fig. S1). Then we analyzed the conformational flexibility of the complex by calculating the root mean square fluctuation (RMSF) for the last 800 ps of the MD simulation (Fig. 1). Obviously, the protein–peptide interaction stabilizes the conformation of the residues close to, particularly the core region of, the interaction interface, as almost all these residues have RMSF values <3 Å (Fig. 1). We also found that the two monomers of the RIIα D/D domain have quite different conformational dynamics indicated by their different RMSF profiles. Overall, monomer 2 is more stable than monomer 1 as the number of residues is 10 and 16 with the criteria of <2.5 Å RMSF for the two monomers, respectively. Particularly, the N terminus of monomer 2 is more stable than that of monomer 1. These observations suggested that AKAP10 might form stronger interaction with monomer 2 than with monomer 1.

We studied the energetic characteristics of the protein–peptide interaction by analyzing the binding free energy components: electrostatic (ΔE_{ele}), van der Waals (ΔE_{vdw}), polar solvation (ΔG_{PB}) and nonpolar solvation (ΔG_{SA}) energies. We observed that the major contribution to the binding is the van der Waals energy (−85.5 kcal/mol). The electrostatic term slightly opposes the binding (6.2 kcal/mol), while the polar solvation term significantly opposes the binding (22.46 kcal/mol). As shown by the contribution of individual peptide residues to the binding free energy using the MM/GBSA decomposition analysis (Fig. 2 and Supplementary Table S2), we found that seven peptide residues (A_{P5}, I_{P8}, A_{P9}, I_{P12}, V_{P13}, V_{P16} and M_{P17}) form relatively stronger interactions than the other residues as their side chains pointing toward the protein (Fig. 2a). These residues also contribute significant van der Waals (ΔE_{vdw}) but not electrostatic/polar ($\Delta E_{es} + \Delta G_{PB}$) interaction energy to binding (Fig. 2b and 2c), which confirms that the hydrophobic interaction is dominant for the AKAP peptide interacting with RIIα. The AKAP10 peptide has a Trp at position P6 (W_{P6}). This residue is not buried in the binding interface but its large hydrophobic side chain can still form effective van der Waals interactions with the protein (−3.12 kcal/mol).

We also evaluated the contribution of each residue in the AKAP10 peptide to the binding free energy with the RIIα D/D domain, and the total contributions and the individual energetic terms of the important residue–residue pairs between AKAP10 and the RIIα D/D domain are listed in Supplementary Table S2. It is apparent that the AKAP residues at positions P5, P6, P8, P9, P12, P13, P16 and P17 have effective contacts with protein residues including Ile3', Ile5(Ile5'), Leu9(Leu9'), Thr10(Thr10'), Leu13(Leu13'), Gln14(Gln14'), Thr17(Thr17'), Val18(Val18'), Thr21(Thr21') and Arg22. The energetic analysis confirmed that the AKAP10 peptide binding is indeed asymmetric to the two monomers of the RIIα D/D domain. The side chains of Ile3' and Ile5' of monomer 2 pointed toward the peptide, making extensive hydrophobic contacts with the residues at positions P4, P5, P8 and P12. In contrast, the N-terminus of monomer 1 is disordered, and the only observed contact is between Ile5 and the peptide residue at P17. Also, the van der Waals interactions are dominant for almost all the residue–residue interactions.

3.2 The binding preferences of residues at the six important positions

Because the residues at positions P5, P8, P9, P12, P13 and P16 form strong interactions with the protein (Fig. 2a) and they are also well conserved (Fig. 3), we studied the preference of amino acids at these positions using the VM method (see Section 2).

Figure 1 shows that Ile_{P8}, Ala_{P9}, Ile_{P12} and Val_{P13} in the peptide are the residues that have the least conformational fluctuation. These residues are in the center of the hydrophobic ridge of the peptide that docks to the hydrophobic pocket of the RIIα D/D domain (Fig. 4). Ile_{P8} forms relatively strong van der Waals interaction with Leu21 (−1.24 kcal/mol), Ile3' (−1.50 kcal/mol), and Ile5'

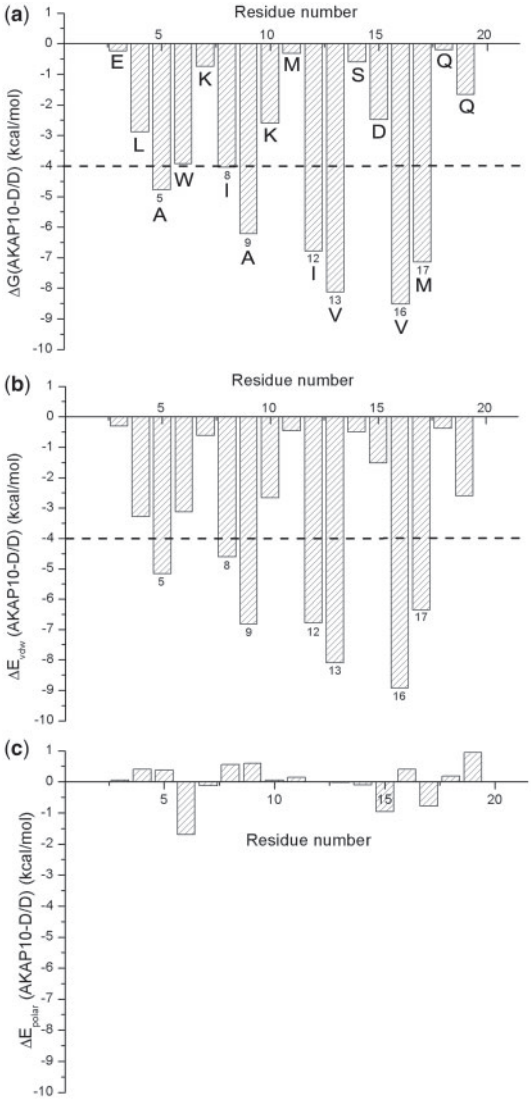

Fig. 2. The interaction energy spectrum for the core residues of the AKAP10 peptide obtained by the MM/GBSA decomposition analysis. (**a**) Total energy, (**b**) van der Waals energy and (**c**) polar energy. The residues with interaction energy less than $-4.0 kcal/mol$ are labeled (the amino acid one-letter codes are labeled in Fig. 2a).

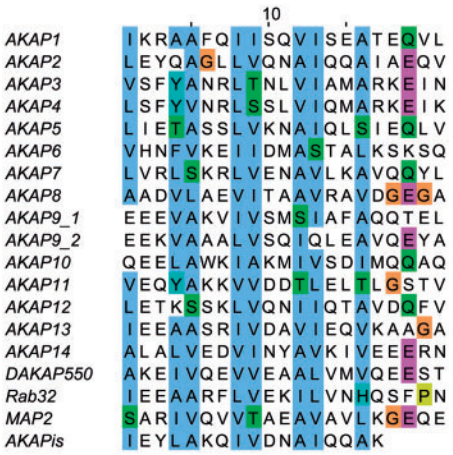

Fig. 3. Alignment of the RII-binding domains of the available AKAPs. Conserved amino acid residues at each position are colored in blue. All but AKAP14 (Kultgen *et al*., 2002) and AKAPis peptides were taken from Hundsrucker *et al*. (2006). The two binding peptides of AKAP9 are labeled as AKAP9_1 and AKAP9_2. AKAPis is a high-affinity peptide obtained from bioinformatics and peptide-array screening (Alto *et al*., 2003). DAKAP550 is a AKAP found in Drosophila, not in human (Han *et al*., 1997).

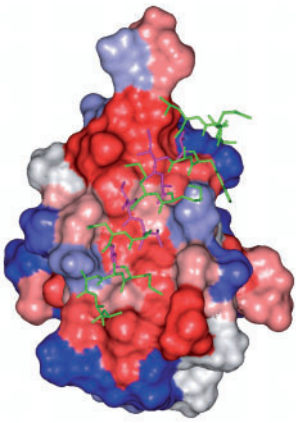

Fig. 4. The hydrophobic binding interface of the RIIα D/D domain. The solvent accessible surface of the RIIα D/D domain is colored by the hydrophobicity of the residues. High hydrophobicity is colored in red and low hydrophobicity in blue. The peptide is shown as a stick model. The picture was generated by Discovery Studio molecular simulation package (2009).

(−0.86 kcal/mol) in the protein (Supplementary Table S2). The VM analysis shows that several other residues, including Phe, Ala, Leu and Tyr, are also favorable at this position (Fig. 5b). Apparently, all residues form significant van der Waals interaction (−85.5 kcal/mol for Ile, −81.2 kcal/mol for Leu, −83.8 kcal/mol for Phe and −78.8 kcal/mol for Tyr) with the protein. Our calculations are consistent with the peptide array experiments, in which Ile is the most favorable residue at this position, and Leu and Phe are also highly favorable (Burns-Hamuro *et al*., 2003). The favorable Ala predicted by our calculation at this position was not observed in the array experiment. The small size of Ala might account for the discrepancy.

Residue at P9 is critical to the peptide binding. Because the binding pocket occupied by the residue at P9 is small large residues are extremely unfavorable at this position (Fig. 5c). The four most favorable residues at this position are Ala (−31.9 kcal/mol), Ser (−25.0 kcal/mol), Ile (−23.9 kcal/mol) and Val (−22.2 kcal/mol). The peptide array experiments found that Ala, Ile and Val are strongly favorable at this position too (Burns-Hamuro *et al*., 2003).

At position P12, Ile, Val and Ala are preferred in our predictions (Fig. 5d), which is in good agreement with the array experiments (Burns-Hamuro *et al*., 2003). Our analysis showed that the van der Waals interactions, the polar contribution to the binding free energy and conformational entropy all play important roles. For example, although Phe at P12 forms strong van der Waals interaction with

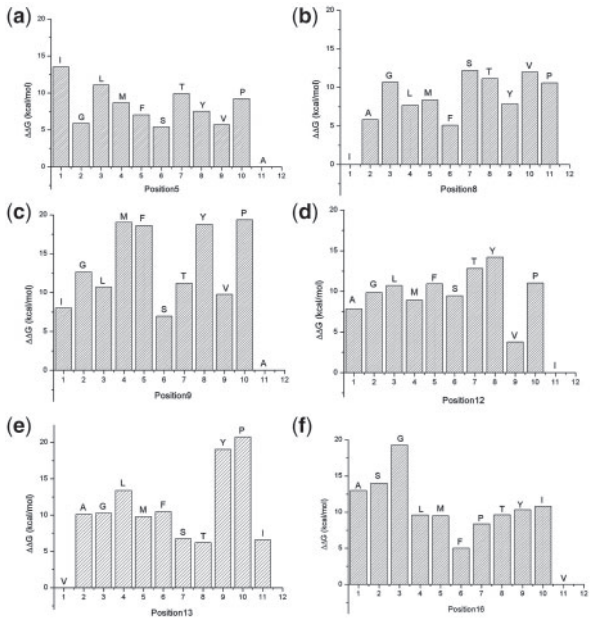

Fig. 5. The binding free energy difference (ΔΔG) between the peptide in the template and the mutated peptide mutated at positions (**a**) P5, (**b**) P8, (**c**) P9, (**d**) P12, (**e**) P13 and (**f**) P16. The preference of amino acids at each position can be determined based on ΔΔG.

the protein (−82.0 kcal/mol), the poor polar interaction and the conformational entropy loss make it unfavorable at this position.

Four residues were predicted to be highly preferred at P13, including Val (−31.9 kcal/mol), Thr (−25.8 kcal/mol), Ile (−25.4 kcal/mol) and Ser (−25.2 kcal/mol), among which Val and Ile were also found favorable at this position by the peptide array experiments (Burns-Hamuro *et al*., 2003). The nonpolar interaction energy $\Delta G_{\text{nonpolar}}$ (the sum of the van der Waals and the nonpolar part of the desovlation free energy) correlates well with the binding free energy ΔG_{bind} ($r = 0.65$), while the polar interaction energy ΔG_{polar} (the sum of the electrostatic and the polar part of the desovlation free energy) does not ($r = 0.37$). This suggests that the nonpolar interaction is the dominant factor for the amino acid preference at this position.

The peptide array experiments showed that positions P5 and P16 are more tolerable to substitutions than positions P8, P9, P12 and P13 (Burns-Hamuro *et al*., 2003). Due to this ambiguity and the conformational flexibility of these two positions, not all favored residues observed in the array experiments, e.g. Ala at P16, have favorable binding energy in the calculation (Fig. 5a and 5f). Finally, we observed that proline was not favorable at almost all positions, which is consistent with the low helix-forming propensity of proline (Macarthur and Thornton, 1991).

3.3 The prediction of the binding peptides for the RIIα D/D domain using a hierarchical approach

Our aim is to identify peptides binding to the RIIα D/D domain in the human proteome, which is a very difficult task. The major hurdle is how to remove false positives. We employed a hierarchical approach to address this issue (Fig. 6). It should be noted that for the top 10 000

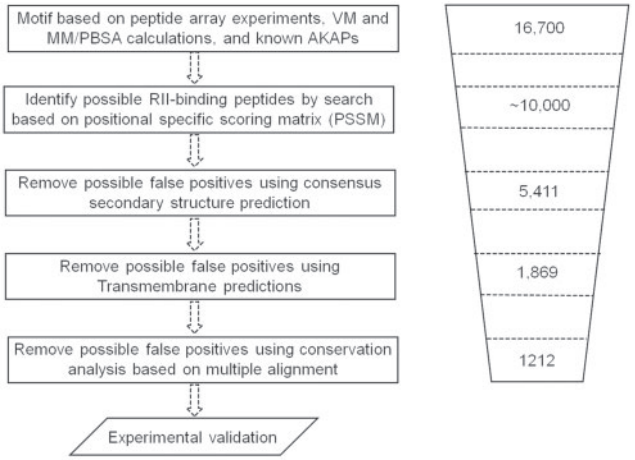

Fig. 6. The hierarchical approach to identify the putative binding peptides of the RIIα D/D domain.

XXXX(A/F/G/I/L/M/S/T/V/Y)XX(F/I/L/V)(A/I/L/S/T/V)XX(A/I/L/S/T/V)(I/L/S/T/V/Q)XX(A/F/G/I/L/S/T/V)XXXXX

Motif1

(A/E/G/I/L/M/Q/S/T/V)xxx(A/F/G/I/L/M/S/T/V/Y)xx(F/I/L/V)(A/I/L/S/T/V)xx(A/I/L/S/T/V/Q)(I/L/S/T/V)XX(A/F/G/I/L/S/T/V)xxxXX

Motif2

Fig. 7. The motifs used in database screening. Little x represents nonproline residues and capital X represents any residue.

peptides ranked by PSSM, inverting two or more steps of filters do not change the final results.

3.3.1 Database screening based on a PSSM Because the RIIα-binding peptides do not show a consensus sequence motif, we used a degenerated pattern (Motif 1 in Fig. 7) based on the known AKAP peptide sequences, the peptide array experiments and the VM calculations. The definitions of the degenerated pattern are described in Part 1 in the supporting materials. When Motif 1 was used to scan the SWISSPROT database, we found 802 970 matches in all species, among which 75 256 were human peptides. After add two constrains to Motif 1, Motif 2 was obtained and the matches to Motif 2 in the SWISSPROT database included about 16 700 human peptides.

Although pattern match can significantly reduce the searching space of the RII-binding peptides, it does not rank the peptides to suggest which peptides are more likely to be true binders. Therefore, we devised a PSSM to rank the 16 700 human 21mer peptides that match to Motif 2 (Boeckmann *et al*., 2003), and the range of the PSSM score is between 6.0 and 94.8. The element in the PSSM was based on the VM calculations, and it is the difference (ΔΔG) between the binding free energy of the mutated peptide and that of the template peptide.

The ranks for the known human RIIα-binding peptides in Figure 3 are shown in Table S3. Four AKAP peptides, AKAP10 (33), AKAP13 (85), AKAP1 (521) and AKAP9_1(894), are ranked in the top 1000. Except AKAP11 and Rab32, the other binders can be found in the top 10 000. Therefore, we kept the top 15 000 peptides for further analysis, which included all known AKAP peptides but Rab32. The residue at P16 of the Rab32 peptide is His, and His at

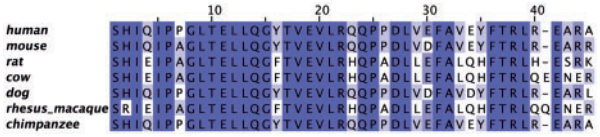

Fig. 8. The alignment of the PKA D/D domains across seven species.

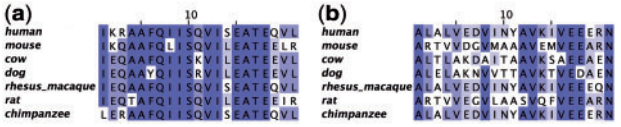

Fig. 9. The sequence alignments across seven species for the binding peptides of (**a**) AKAP1 and (**b**) AKAP14.

P16 is absent from the predefined motifs, so the motif search cannot identify Rab32.

3.3.2 Secondary structure prediction One of the unique features of the AKAP peptides is the amphipathic α-helix. Therefore, secondary structure prediction was used as a filter to remove false positives. A consensus score was computed by averaging the two highest percentages given by the three methods. Among the 18 known human AKAP peptides shown in Figure 3, 13 were predicted to have 100% α-helices, and 4 have >90% α-helices.

Because 17 of the 18 known AKAP peptides except the AKAP12 peptide were predicted to be >90% α-helices, 90% was used as a threshold to determine whether a peptide forms α-helix. Using this criterion, 5411 out of the 15 000 (36.1%) peptides were predicted to form α-helices. After we applied the consensus secondary structure prediction as a filter to remove false positives, the ranks of the known RIIα-binding peptides were improved significantly (Supplementary Table S3). For the 17 known human AKAP peptides shown in Supplementary Table S3, 14 could be found in the top 2000 predicted peptides.

3.3.3 TM prediction It is obvious that a TM helix cannot form direct interaction with PKA. Indeed, all the 17 known AKAP peptides shown in Supplementary Table S3 are not predicted to be located in TM regions. Therefore, TM prediction was used as a filter to remove false positives, which further improved the ranks of the known RIIα-binding peptides significantly (Supplementary Table S3) and the number of candidate AKAP peptides in the human proteome was decreased to 1869. In other words, 65.5% of the 5411 candidate peptides were removed. For the 17 known AKAP peptides in Supplementary Table S3, 14 could be found in the top 700.

3.3.4 Conservation analysis Alignment of RIIα D/D domain across 7 species (Fig. 8) showed that the D/D domain is highly conserved and the variation positions are located far from the binding site. Because of the functional constraint, the peptides binding to the RIIα D/D domain are also expected to be conserved across species. When all the 21 positions in the AKAP peptide were considered in the conservation analysis, only two of the 15 known AKAP peptides, AKAP1 and AKAP14, have the average similarities lower than 0.90 (0.88 for AKAP1 and 0.66 for AKAP14). Examination of the multiple sequence alignments of these two peptides showed that the variation positions do not contribute directly to the peptide binding, i.e. positions P2, P3, P11, P14, P19, and P21 in AKAP1, and positions P2, P3, P6, P7, P10, P11, P14, P15 and P19 in AKAP14 (Fig. 9). Consistently, if we only consider the eight important positions in the peptide that interact directly with the protein, including P1, P4, P5, P8, P9, P12, P13 and P16, all the 15 known human AKAP peptides are well conserved. Twelve are completely conserved across seven species, i.e. the average similarity is 1.00; the rest AKAP peptides, AKAP9_1, AKAP4 and AKAP14, also have high average similarities of 0.97, 0.90 and 0.89, respectively. Therefore, we chose to only consider the eight important positions in the conservation analysis and used an average similarity of 0.88 as a cutoff to remove false positives. About 35.2% peptides were removed by the conservation filter (the number of peptides in the final list is 1212).

3.3.5 Analysis of the predictions by the hierarchical strategy The secondary structure, TM and conservation filters reduced the number of the potential RIIα-binding peptides from 15 000 to 1212. The ranks given by the PSSM showed that 14 known AKAP peptides present in the top 450, ten in the top 300 and three in the top 50. It is obvious that the PSSM can capture the important binding characteristics of the AKAP peptides. The only exception is AKAP11 with a low rank (1141). The basic reason for the low rank of AKAP11 is that Thr at P12 is not very energetically favorable ($\Delta\Delta G = 12.8$ kcal/mol), therefore the rank given by the PSSM score for the AKAP11 peptide is not so high. Another possible explanation is that the binding capability of AKAP is not only completely determined by the known AKAP peptide but also by the flanking residues around this AKAP peptide. Certainly the contribution of the flanking residues cannot be estimated by the approach used in this study. The binding peptides of AKAP12 and Rab32 shown in Figure 3 could not be found because the Rab32 peptide did not match the searching motif pattern and the AKAP12 peptide was not predicted to form a helix.

Most known AKAPs contain multiple peptides that are ranked in the top 15 000 using the PSSM and the highest ranked one is often the documented binding peptides (Hundsrucker *et al*., 2006) (Supplementary Table S3). We also found new peptides ranked highest in several of these AKAPs including AKAP8, AKAP11 and MAP2 (Table S3). For example, AADVLAEVITAAVRAVDGEGA (rank: 311) was documented as the binding peptide of AKAP8 (Hundsrucker *et al*., 2006). Our analysis ranked peptide EEVAADVLAEVITAAVRAVDG (rank: 10) much higher. Coghlan *et al*. (1994) showed that solid-phase RIIα binding to the rat AKAP8 was inhibited by a synthetic peptide of EVAAEVLAEVITAAVKAV , which is a segment of the higher ranked peptide in our prediction. This evidence supports that our analysis can identify new peptides binding to RIIα. Similarly, two binding peptides were documented for MAP2: SARIVQVVTAEAVAVLKGEQE (rank: 604) (Hundsrucker *et al*., 2006) and AEEVSARIVQVVTAEAVAVLK (rank: 146) (Scholten *et al*., 2006). Our analysis strongly supports the latter peptide to be a tighter binding sequence and this prediction is waiting for experimental validation.

The PKA binding region in AKAP14 proposed by Hundsrucker *et al*. (2006) is TQDKNYEDELTQVALALVEDVINYA, in which a segment YEDELTQVALALVEDVINYAV was found by the PSSM search. However, this peptide is not conserved and thus was removed

by the conservation filter. Our search found another binding peptide in AKAP14: ALA**L**VED**V**INYAVKIVEEERN (rank: 131) and its first 12 residues are the same as those at the C-terminus of the binding peptide proposed by Hundsrucker *et al*. (2006). Kultgen *et al*. (2002) reported that mutations of L43 and V47 (the bold residues in the above sequence) to proline completely abolished the RII-binding, which strongly supported the predicted peptide bound to the RIIα D/D domain.

A known AKAP, WAVE-1 (Westphal *et al*., 2000), was not included in the initial alignment. Two putative RIIα-binding peptides, TNISLANIIRQLSSLSKYAED and VISDARSVLLEA**I**RKG**I**QLRK (rank: 709), were found by the PSSM search in the top 15 000; but only the latter peptide was predicted to form α-helix and found conserved across the seven species. Our prediction result was supported by the mutation experiments, in which substitution of isoleucines 505 and 509 by proline destroyed RII interaction (Westphal *et al*., 2000). Interestingly, although WAVE-1 homologs, WAVE-2, WAVE-3 and WAVE-4 show significant sequence similarity to the putative RIIα-binding domain of WAVE-1, the predicted AKAP peptide in WAVE-1 is not conserved in the other three WAVE proteins. WAVE-2, WAVE-3 and WAVE-4 can be hypothesized to not interact with the RIIα D/D domain. The biological implication of such interaction specificity needs further experimental investigation.

In the mouse protein MyRIP, Goehring *et al*. (2007) recently reported that two peptides VALRVAEEAIEEAISKA (position 193–209) and LTEELAGTILQRIIRKQ (position 232–248) were RII-binding regions using Solid phase peptide array . They further demonstrated that the PKA-anchoring site on MyRIP is located between residues 194 and 210 while residues 232–248 regulate the RII binding negatively because the mutations at positions 236 and 245 increase the RII binding of MyRIP when compared with the wild-type protein. A peptide in the human MyRIP (TLAVALRVAEEAIEEAISKAE) was ranked 388 in our predictions, which is consistent with the experimental observation in mouse (Goehring *et al*., 2007). Given that the binding peptide is highly conserved, we conclude that the MyRIP proteins in all the seven species may be AKAPs.

In Figure 3, DAKAP550 is a AKAP discovered in Drosophila (Han *et al*., 1997). One homology of DAKAP550, LRBA (Beige-like protein, BLP), was found in the human proteome. It is interesting to find that one peptide in LRBA is also included in our prediction: SIEASVTFLQRLISLVDVLIF (rank: 280). Given the high sequence similarity between DAKAP550 and LRBA, our prediction results suggest that human protein LRBA uncharacterized previously may be involved in PKA/AKAP interaction.

4 CONCLUSIONS

We have conducted systematic studies using MD simulations and free energy calculations to characterize the binding specificity between the human PKA RIIα regulatory subunit and its binding peptides. We evaluated the binding preferences at several peptide positions that interact directly to RIIα quantitatively. Our calculation results were in good agreement with the qualitative measurements by the peptide array experiments.

It is very challenging to predict the interacting peptides to RIIα. To tackle this challenge, we employed a hierarchical procedure to identify the RIIα-binding peptides from the human proteome successfully. Our goal is to identify a list of putative binding peptides that are ranked based on their binding with the PKA regulatory subunit. Even without a gold standard to judge the accuracy of our predictions, the supporting evidence for numerous predicted peptides was indeed found in the literature. Given the difficulty of identifying AKAP peptides, our study provides a useful candidate list that can narrow down the searching space for experimentalists.

ACKNOWLEDGEMENTS

The authors are grateful to the associate editor and anonymous referees for comments and helping to improve the earlier version. The authors thank Drs Susan Taylor and William A. McLaughlin for their critical discussions.

Funding: T.J. and Y.Y. are supported by Natural Science Foundation of China (No. 20973121) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). W.W. is supported by National Institute of Health Grant (No. R01GM085188).

Conflict of Interest: none declared.

REFERENCES

Discovery Studio 2.5 Guide (2009) Accelrys Inc., San Diego.

Alto,N.M. and Scott,J.D. (2004) The role of A-kinase anchoring proteins in cAMP-mediated signal transduction pathways. *Cell Biochem. Biophys.*, 201–207.

Alto,N.M. *et al*. (2003) Bioinformatic design of A-kinase anchoring protein-in silico: a potent and selective peptide antagonist of type II protein kinase A anchoring. *Proc. Natl Acad. Sci. USA*, **100**, 4445–4450.

Altschul,S.F. (1991) Amino-acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.*, **219**, 555–565.

Altschul,S.F. *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Appert-Collin,A. *et al*. (2006) Review - regulation of G protein-coupled receptor signaling by A-kinase anchoring proteins. *J. Recept. Sig. Transd. Res.*, **26**, 631–646.

Apweiler,R. *et al*. (2004) Protein sequence databases. *Curr. Opin. Chem. Biol.*, **8**, 76–80.

Boeckmann,B. *et al*. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.

Brannetti,B. *et al*. (2000) SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. *J. Mol. Biol.*, **298**, 313–328.

Burns-Hamuro,L.L. *et al*. (2003) Designing isoform-specific peptide disruptors of protein kinase A localization. *Proc. Natl Acad. Sci. USA*, **100**, 4072–4077.

Carnegie,G.K. and Scott,J.D. (2003) A-kinase anchoring proteins and neuronal signaling mechanisms. *Gene Dev.*, **17**, 1557–1568.

Case,D.A. *et al*. (2005) The amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.

Coghlan,V.M. *et al*. (1994) Cloning and characterization of Akap-95, a nuclear-protein that associates with the regulatory subunit of type-II camp-dependent protein-kinase. *Faseb J.*, **8**, A1226–A1226.

Cuff,J.A. and Barton,G.J. (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*, **40**, 502–511.

Darden,T. *et al*. (1993) Particle Mesh Ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

Duan,Y. *et al*. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.

Ferraro,E. *et al*. (2006) A novel structure-based encoding for machine-learning applied to the inference of SH3 domain specificity. *Bioinformatics*, **22**, 2333–2339.

Fong,J.H. *et al*. (2004) Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.*, **5**, R11.

Frishman,D. and Argos,P. (1997) Seventy-five percent accuracy in protein secondary structure prediction. *Proteins*, **27**, 329–335.

Goehring,A.S. *et al*. (2007) MyRIP anchors protein kinase a to the exocyst complex. *J. Biol. Chem.*, **282**, 33155–33167.

Han,J.D. *et al*. (1997) Molecular characterization of a novel A kinase anchor protein from Drosophila melanogaster. *J. Biol. Chem.*, **272**, 26611–26619.

Hou,T.J. *et al*. (2006a) Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain. *Plos Comput. Biol.*, **2**, 46–55.

Hou,T.J. *et al*. (2006b) Prediction of binding affinities between the human amphiphysin-1 SH3 domain and its peptide ligands using homology modeling, molecular dynamics and molecular field analysis. *J. Proteome Res.*, **5**, 32–43.

Hou,T.J. *et al*. (2008) Characterization of domain-peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. *J. Mol. Biol.*, **376**, 1201–1214.

Hou,T.J. *et al*. (2009) Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. *Mol. Cell Proteomics*, **8**, 639–649.

Hundsrucker,C. *et al*. (2006) High-affinity AKAP7 delta-protein kinase A interaction yields novel protein kinase A-anchoring disruptor peptides. *Biochem. J.*, **396**, 297–306.

Kim,C. *et al*. (2005) Crystal structure of a complex between the catalytic and regulatory (RI alpha) subunits of PKA. *Science*, **307**, 690–696.

Kinderman,F.S. *et al*. (2006) A dynamic mechanism for AKAP binding to RII isoforms of cAMP-dependent protein kinase. *Mol. Cell*, **24**, 397–408.

Kollman,P.A. *et al*. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts Chem. Res.*, **33**, 889–897.

Krogh,A. *et al*. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.

Kultgen,P.L. *et al*. (2002) Characterization of an A-kinase anchoring protein in human ciliary axonemes. *Mol. Biol. Cell*, **13**, 4156–4166.

Langeberg,L.K. and Scott,J.D. (2005) A-kinase-anchoring proteins. *J. Cell Sci.*, **118**, 3217–3219.

Macarthur,M.W. and Thornton,J.M. (1991) Influence of proline residues on protein conformation. *J. Mol. Biol.*, **218**, 397–412.

McLaughlin,W.A. *et al*. (2006) Prediction of binding sites of peptide recognition domains: an application on Grb2 and SAP SH2 domains. *J. Mol. Biol.*, **357**, 1322–1334.

Obenauer,J.C. *et al*. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.

Onufriev,A. *et al*. (2000) Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.

Rost,B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Method Enzymol*, **266**, 525–539.

Ryckaert,J.P. *et al*. (1977) Numerical-integration of Cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J. Comput. Phys.*, **23**, 327–341.

Schleinkofer,K. *et al*. (2004) Comparative structural and energetic analysis of WW domain-peptide interactions. *J. Mol. Biol.*, **344**, 865–881.

Scholten,A. *et al*. (2006) Analysis of the cGMP/cAMP interactome using a chemical proteomics approach in mammalian heart tissue validates sphingosine kinase type 1-interacting protein as a genuine and highly abundant AKAP. *J. Proteome. Res.*, **5**, 1435–1447.

Thompson,J.D. *et al*. (1994) Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Wang,J.M. *et al*. (2006) Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr. Comput. Aided Drug Des.*, **2**, 287–306.

Wang,W. *et al*. (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Bioph. Biom.*, **30**, 211–243.

Westphal,R.S. *et al*. (2000) Scar/WAVE-1, a Wiskott-Aldrich syndrome protein, assembles an actin-associated multi-kinase scaffold. *EMBO J.*, **19**, 4589–4600.

Xiang,Z.X. and Honig,B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.*, **311**, 421–430.