Interactions between CD44 and HA$_{16}$: An investigation on multiple binding modes of the complex by using molecular dynamics simulation studies

Dipankar Paul, Santanu Santra and Madhurima Jana*

Molecular Simulation Laboratory, Department of Chemistry, National Institute of Technology, Rourkela-769 008, Odisha, India

E-mail: janam@nitrkl.ac.in

Hyaluronan (HA), a polyanionic polysaccharide acts as a signal transmitter molecule and regulates various cell functions while binding with a receptor protein CD44. X-ray crystallographic study has revealed one single binding mode however; multiple binding modes of CD44-HA complex have been identified by several researchers. In this work, we characterize the three different binding modes of CD44-HA$_{16}$ namely, crystallographic, parallel, and upright modes by atomistic Molecular Dynamics simulations and Molecular Mechanics-Poisson Boltzmann-solvent accessible surface area (MM/PBSA) calculations. Attempts have been made to study the conformational flexibilities and the relative binding affinities of the three complexes. The interactions that contribute more to the overall binding free energies of the complexes have been investigated in detail. Further, the residues that are involved in the recognition process were identified. Our study infers that the recognition process is generally governed by the apolar interactions. It is apparent from our calculations that the polar contribution is unfavorable for the parallel and upright modes, however crystallographic binding mode is noted to be facilitated by the favorable polar interaction and the direct hydrogen bonding network that forms between CD44 and HA$_{16}$.

Keywords: CD44, hyaluronic acid, molecular dynamics, MMPBSA, binding interaction.

1. Introduction

Hyaluronic acid (HA) commonly known as hyaluronan is a negatively charged unbranched natural polymer composed of repeating disaccharide units of N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcUA), (β-1,3-GlcNAc-β-1,4-GlcUA)$_n$. HA acts as a molecular lubricant, and plays important role in embryonic development, wound healing and in processes such as leukocyte trafficking etc.[1]. It is a pervasive component of the extracellular matrix. HA is the major ligand for CD44, a transmembrane receptor protein that is highly expressed in many cancers and regulates metastasis. It is also known as a marker for stem cells. HA can activate cytoskeleton and matrix metalloproteinases signaling progression upon binding with CD44[2]. Molecular weight of HA has crucial role in the biological function[3] such as; in tumorigenesis, antiangiogenic and anti-inflammatory responses high molecular weight HA is involved whereas low molecular weight HA has shown its efficiency in promoting cell motility etc. Considering the fact that CD44-HA interactions play an important role in biological recognition processes several interesting aspects of the complexes have been explored by using various experimental and simulation methods. The human CD44 consists of 723 residues of which 158 residues known as HA binding domain (HABD)[4]. Banerji *et al.*[5] have resolved the structure of HA bound to CD44, where two conformations named as 'A' and 'B' were identified. They proposed that 'B' form establishes more intimate contact with HA through a shallow binding groove of the link module of HABD. Further, other experimental studies including NMR experiment showed that there exist two fundamental conformational forms of CD44, of which one is very similar to the structure proposed by Banerji *et al.* named as 'ordered' conformation and the other that differs in the beta-sheet region was named as 'partially disordered' conformations[6,7]. Although both the structures were proposed to bind HA, computer simulation study had shown preference of partially disordered form over the ordered conformation to bind HA[8]. The study further inferred that the conformational freedom

gained by the C-terminal residues allows the 'partially disordered' conformer to bind the ligand. Apart from this, Teriete et al.[4] proposed another mode of binding known as 'upright' mode where HA occupies a region perpendicular to the binding groove. Martinez-Seara and coworkers[9] suggested one more binding mode of CD44-HA, where the sugar rings of HA lie on the 'lower' part of the binding groove known as 'parallel' mode. Their studies have shown that the crystallographic mode is not the most easily accessible binding mode although it may form strong complex.

Most of the aforementioned studies have identified multiple binding modes of CD44-HA complex however, till now X-ray crystallographic study has revealed only one single binding mode of the complex. Hence it will be of great interest to look after the comparative binding affinities of several binding modes of CD44-HA complex. As mentioned earlier, in a recent work, Martinez-Seara and coworkers[9] have studied the multiple binding modes of the complexes. Driven by their studies in this work, we characterize the three different binding modes of CD44-HA namely, crystallographic, parallel, and upright modes by performing atomistic molecular dynamics simulations and Molecular Mechanics-Poisson Boltzmann-solvent accessible surface area (MM/PBSA) method. Attempts have been made to study the conformational flexibilities of the three complexes and to quantify the binding free energy of the complexes to provide an estimation how these three modes of binding differ from each other. Further, the details of interactions that might play role in the complexation process have been thoroughly investigated. In Fig. 1 we display the aligned structures of the three different binding modes of CD44-HA complex. The orientation of HA shows its relative initial position on the CD44 surface. The HABD of CD44 as shown in the figure contains 150 amino acid residues. For convenience, these residues are numbered as 1 to 150 in our study, which actually correspond to the residues 20 to 170 of the CD44. The regular secondary structural segments of the protein consist of two α-helices, one 3_{10}-helix, and ten β-strands connected by coils. The studied HA oligomer is HA_{16}, the suffix denotes the number of monosaccharide units present. The starting and end residues of HA for each binding mode are also marked in Fig. 1 for convenience. The rest of the article is organized as follows. In Section 2, we describe in brief the setup of the systems and methodology employed. The results are presented
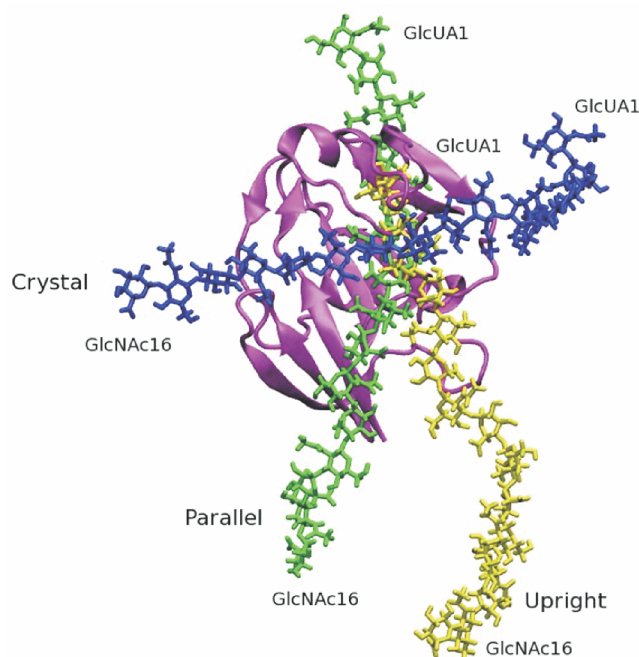

Fig. 1. The three binding modes CD44-HA$_{16}$ complex shown. The whole CD44 is shown in magenta, and the HA in crystallographic, parallel and upright binding modes are shown in blue, green and yellow respectively. The starting and ending residues of HA moiety are marked for convenience.

in detail and discussed in Section 3. The important findings and the conclusions reached from this study are summarized in Section 4.

2. Methodology

2.1. MD simulation

In this work we have carried out atomistic molecular dynamics (MD) simulations of CD44-HA$_{16}$ complex considering its three well-defined binding modes[5,9] designated as C1, C2, and C3 by using the AMBER code[10]. In rest of our study, the crystallographic binding mode as reported by Banerji et al.[5] is designated as C1 complex, whereas the parallel and upright binding modes as described by Martinez-Seara and coworkers[9] and Teriete et al.[4] are designated as C2 complex, and C3 complex respectively. The initial coordinates of all the three modes of the CD44-HA$_{16}$ complex were taken from the literature[9]. The two terminal residues of CD44 were taken as standard ammonium and carboxylate ionic forms, while the GlcUA1 and GlcNAc16 monomers of the carbohydrate were capped by hydroxylating the terminal glycosidic oxygen atoms. Each of the complexes was then immersed

in a separate large cubic cell of edge length 100 Å of well-equilibrated water molecules. To avoid any unfavorable contact, the insertion process was carried out by carefully removing all the water molecules that were present within 2 Å from any of the atoms of the complex. The overall charge of the system was neutralized by adding 13 Na$^+$ ions.

To eliminate any initial stress, each system was first minimized using the conjugate gradient energy minimization method as implemented in the AMBER code[10]. The temperature of the system was then gradually increased to the room temperature of 300 K within a short MD run of about 100 ps. This was carried out at a constant pressure (P = 1 atm) under the isothermal-isobaric ensemble (NPT) conditions. It was then followed by 2 ns NPT and 2 ns NVT equilibration run at 300 K. After this step, a long NVT production run was carried out for about 50 ns duration. We have used Langevin dynamics method[11] to control the temperature using collision frequency of 1 ps^{-1} whereas the pressure was controlled by using a Berendsen barostat[12] with a pressure coupling constant of 2.0 ps. A 8 Å cutoff was used for all nonbonded interactions and the SHAKE algorithm[13] in order to constrain all bonds involving hydrogen. All the three simulations were carried out with a time step of 1 fs. The electrostatic interactions were evaluated fully by using the particle-mesh Ewald (PME) method[14] and periodic boundary conditions. We have employed the all-atom AMBER ff12SB force fields and potential parameters for the protein[15], while the parameters for the carbohydrate were taken from the GLYCAM06h[16] force field. The TIP3P model[17] which is consistent with the chosen protein and carbohydrate force fields was employed to model the water molecules. It may be noted that the AMBER force field used in this study has been found satisfactory in reproducing experimental binding affinities of the protein-ligand complexes[18–20]. Although it may be of general interest to estimate the binding affinities by using different other force fields to verify the results, we believe that the general conclusions obtained from our study would not be affected with the variation of the force fields. Presently, we are working on such aspects in our laboratory.

2.2. Binding free energy calculation

The binding free energy of CD44 and HA$_{16}$ was calculated by both MM/PBSA and MM/GBSA methods[21] as implemented in AMBER 12[10]. For each complex we have extracted 6000 frames from the last 40 ns of production trajectories of

the MD simulations and then the average binding free energy for each complex was estimated by using the following expression,

$$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{solv} - T\Delta S_{conf}$$

where ΔG_{MM} is the molecular mechanics energy of the molecule. It includes internal energy and the electrostatics as well as van der Waals interactions. ΔG_{solv} is the free energy of solvation upon binding with HA$_{16}$. It can be written as,

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{SA}$$

where $\Delta G_{PB/GB}$ is the polar contribution calculated from the MM/PBSA or MM/GBSA methodology. In PBSA method the polar solvation term was calculated by using Poisson model whereas in GBSA method the polar contribution of the solvation energy was calculated by using Generalized Born (GB) model[22]. In PBSA method the nonpolar contribution was determined by solvent accessible surface area (SASA) using the MSMS program[23] while for GBSA model the contribution was estimated by using LCPO method[24]. It may be noted that MM/GBSA is one of the faster variants to the MM/PBSA method in which the generalized Born (GB) solvent model is used. We have used both the methods to compare the relative trend of binding affinities of the three different binding modes of CD44-HA$_{16}$. It has been reported that while the performance of PBSA is sensitive to the solute dielectric constant, performance of GB model depends on the GB model used. Since MM-PBSA/GBSA does not include entropy contribution the conformational entropy calculation was performed by using normal-mode analysis as implemented in AMBER12 NMOD program[25,26]. Due to high computational cost in the NMOD calculation total 40 snapshots were extracted from the last 40 ns MD trajectory and they were used for the subsequent calculations. To identify the residues that are responsible for the three binding modes of CD44-HA$_{16}$ complexes, we have further calculated the interactions between the residues of CD44 and HA$_{16}$ by using the MM/PBSA decomposition process[27,28]. The residue wise binding energy consists of van der Waals, electrostatic, and polar solvation contribution.

3. Results and discussion

3.1. Conformational flexibility of the complexes

In Fig. 2 we have shown several configurations of the CD44-HA complexes as taken from different time intervals of the three different simulations. The energy minimized struc-
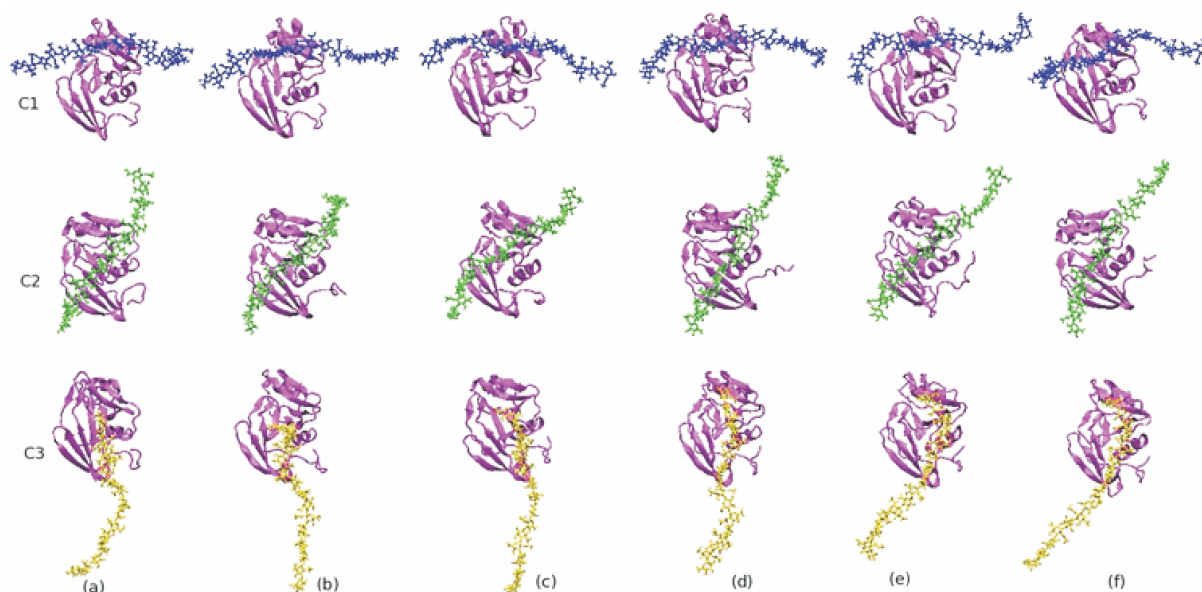
Fig. 2. (a) The energy minimized snapshot of CD44-HA complex and (b-f) the simulated configurations as obtained at 10, 20, 30, 40, and 50 ns of the three separate simulations of C1, C2 and C3 complex in aqueous medium. The coloring scheme is as same as shown in Fig. 1.

tures of the complexes are also included in the figure for comparison. It is apparent from the figure that the overall structural features of the complexes are similar with the corresponding energy minimized structure. Further to visualize the structural deviation of the simulated complexes with respect to their initial structures we have superimposed the backbone N, C, and Cα atoms of CD44 and the non-hydrogen atoms of HA as taken from several configurations of the equilibrated trajectories at a regular interval of 5 ns on the initial starting structure. This is shown in Fig. 3. The figure infers that the backbone of CD44 in C1 and C3 complexes is significantly rigid than in C2 complex whereas HA is noted to be more rigid in C2 relative to C1 and C3 complexes. It is important to note from the Fig. 3 that the part of the residues of HA that are in close contacts with CD44 are relatively more rigid than the rest part of HA. This primarily suggests that the recognition process of CD44-HA complex is associated with the reduction in flexibility of the binding residues of the complexes. However, this requires further investigation. We would address this issue in the forthcoming section.

The differences between the simulated and the initial structures of the complexes were quantified by estimating the root mean square deviations (RMSD). Estimation of

RMSD is a common practice to know the stability of a biomolecule. We have calculated RMSDs for the three complexes as well as for CD44 and HA separately. The time evolution of RMSDs is displayed in Fig. 4. Further, in Table 1 we have listed the time average RMSD values of the complexes along with the individual average RMSD value of CD44 and HA_{16}. The RMSDs were calculated over the entire simulated trajectories by considering the heavy-atoms of CD44-HA. It is evident from Fig. 4(a) that all the three complexes are comparably flexible in nature, however among the three complexes C3 is found to be less rigid. This is further reflected from the average RMSD values as shown in Table 1. The individual contribution of CD44 and HA_{16} on the overall flexibilities of the three complexes can be observed from the time evolution of RMSDs of individual CD44 and HA_{16} molecule in the complexes, as shown in Fig. 4(b) and (c). It may be observed from the figure that the HA-dynamics play important role on the conformational stability of the complexes. We find that among the three complexes, in C1 complex, CD44 is quite rigid whereas the rigidity of HA_{16} is maintained more in C2 complex. The relative trend of such flexibilities of CD44 and HA_{16} is further evident from the average RMSD values as shown in Table 1.
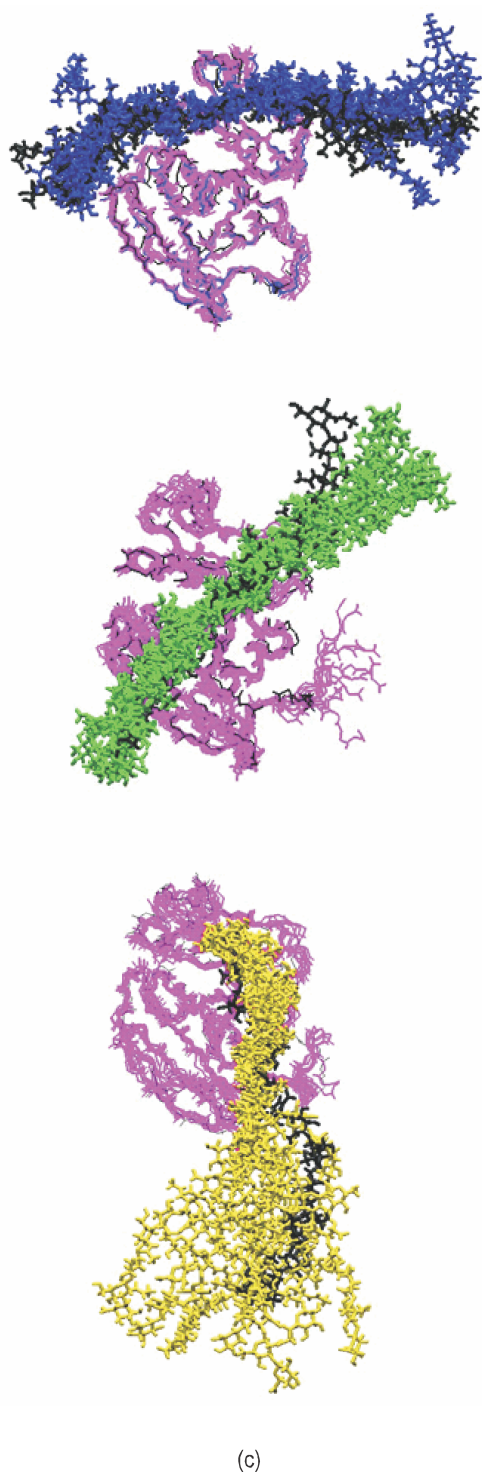
(c)

Fig. 3. Superposition of the backbone atoms of CD44 and the non-hydrogen atoms of HA$_{16}$ obtained from simulated configurations of C1 (a), C2 (b), and C3 (b) complex at a regular interval of 5 ns from the last 45 ns equilibrated trajectories. The initial starting structure of CD44 and HA are shown in black.
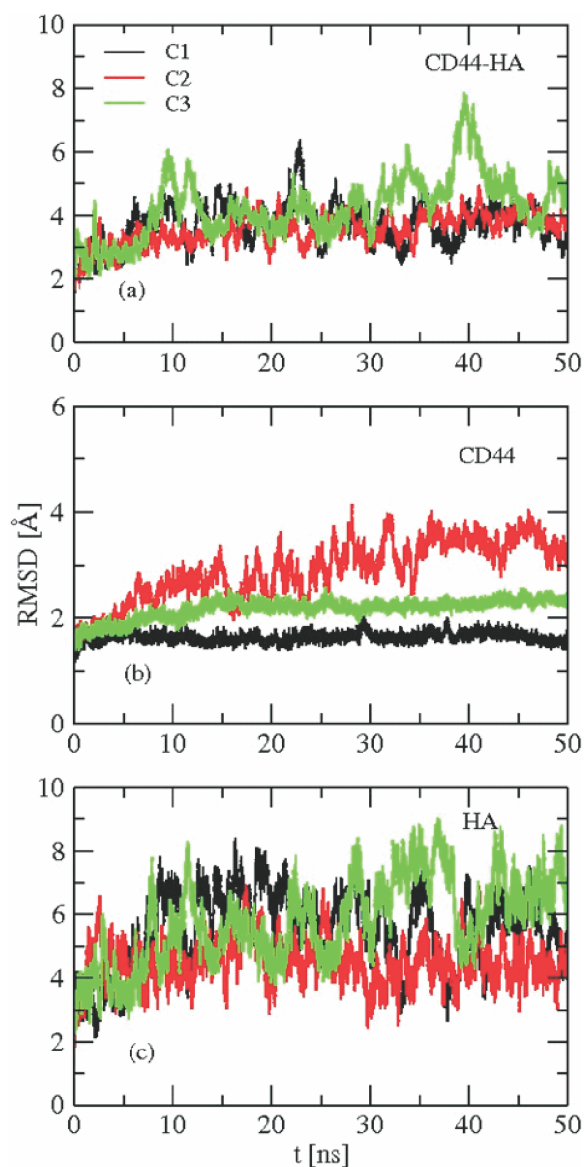

Fig. 4. Time evolution of the RMSDs for all the non-hydrogen atoms of (a) the three complexes as obtained from the entire simulated trajectories. The RMSDs of (b) CD44 and (c) HA$_{16}$ in their complexed forms are also shown.

Table 1. Average RMSD (in Å) values of all the non-hydrogen atoms of C1, C2, and C3 complexes as obtained from the simulations. The corresponding average RMSD values of CD44 and HA$_{16}$ in their complexed forms are also listed

System	< RMSD > (in Å)		
	Whole complex	CD44	HA$_{16}$
C1	3.80	1.63	5.73
C2	3.55	2.89	4.52
C3	4.28	2.17	5.78

The fluctuation of each residue of CD44 and HA_{16} was additionally analyzed by calculating root-mean-square fluctuation (RMSF). It is apparent from Fig. 5(a) that the residues of CD44 is considerably more flexible in C2 relative to that in C1 and C3. The results as shown in Fig. 5(a) infer that mostly the RMSF value of each residue of CD44 in C1 and C3 complexes is in marginal order, indicating no major differences between the mean structure over a dynamic ensemble, except for the residues that lie near to one of the terminus. However, for C2 complex the fluctuation becomes relatively higher in magnitude particularly for the residues that are near to the C-terminus. The RMSF of HA_{16} shows interesting results. It is apparent from Fig. 5(b) that in C1 complex, the residues GlcUA5---GlcUA11 of HA_{16} exhibit lesser flexibility than the other residues of the HA molecule. Therefore, these seven residues of HA_{16} are expected to make prominent contacts with the CD44 while forming C1 complex. On the other hand, considerably lesser flexibility for the residues GlcNAc6---GlcNAc14 and GlcNAc2---GlcNAc6 as observed for C2 and C3 complexes, respec-

tively infers that these residues could contribute more towards binding. In the next section we have analyzed this again by calculating the binding free energy.

3.2. Binding free energy between CD44 and HA $_{16}$

Local conformational flexibilities in CD44 and HA_{16} can affect the binding affinities of the three binding modes of CD44-HA complex. To investigate this, we have calculated the binding free energy of the three complexes by using MM/PBSA and MM/GBSA methods as described earlier. In Table 2 we have listed the binding free energies ($\Delta G_{bind(PBSA/GBSA)}$) and the related components of the three complexes as obtained from our MM/PBSA and MM/GBSA calculations, respectively. It is evident from the Table 2 that CD44 has strong binding affinities towards the ligand HA_{16}, irrespective of the
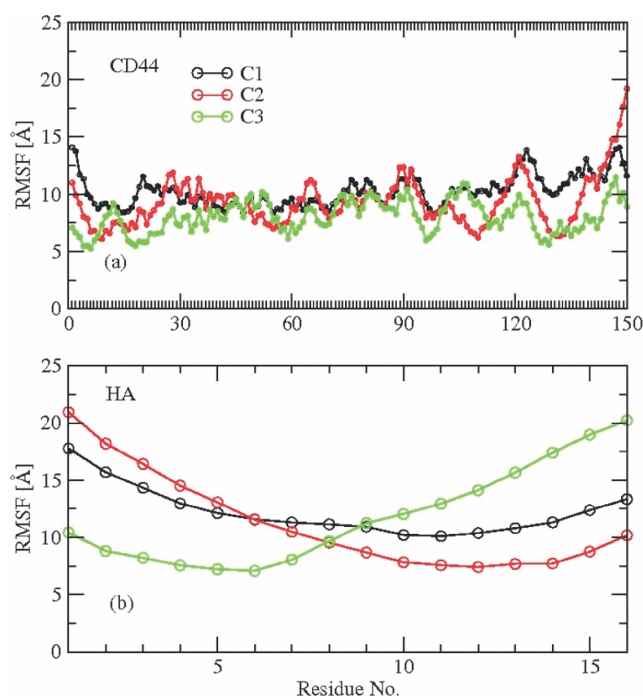
Table 2. The binding free energies of C1, C2 and C3 complexes and the individual energy term as obtained from the MM/PBSA method. The binding free energies as obtained from the MM/GBSA method are also included

Free energy contribution (kcal/mol)	C1	C2	C3
ΔG_{vdw}	−41.47	−75.46	−60.54
ΔG_{EEL}	173.63	114.43	57.18
ΔG_{PB}	−176.34	−107.05	−49.64
ΔG_{SA}	−4.65	−8.23	−7.11
$\Delta G_{MM-PBSA}$	−48.82	−76.31	−60.11
$-T\Delta S$	33.46	56.81	48.84
$\Delta G_{vdw} + \Delta G_{SA}$	−46.12	−83.69	−67.65
$\Delta G_{EEL} + \Delta G_{PB}$	−2.71	7.38	7.54
ΔG_{bind} (PBSA)	−15.36	−19.5	−11.27
ΔG_{bind} (GBSA)	−7.85	−8.13	−7.71

modes of binding. MM/PBSA calculation reveals that the overall binding free energy of C3 complex (−11.27 kcal/mol) is less favorable than the C1 (−15.36 kcal/mol) and C2 (−19.50 kcal/mol) complexes. The values suggest that among the three complexes the binding affinity is most favorable for the parallel mode whereas the crystallographic mode lies in between two. Similar qualitative trend of binding affinities of the three complexes is additionally observed from the MM/GBSA calculations. In this context it may be noted that Vuorio *et al.* have shown that the crystallographic mode forms stronger

Fig. 5. RMSF values of each residue of CD44 (a) and HA_{16} (b) in the three complexes as obtained from the equilibrated trajectories of the three simulations.

complex whereas the most probable mode is the parallel mode[9]. According to Table 2, the van der Waals energy contributes more favorably than electrostatic to the CD44-HA$_{16}$ binding, this is true irrespective of the mode of binding. Both polar (ΔG_{PB}) and nonpolar (ΔG_{SA}) solvation free energies are found to favor the complex formation process. The entropic contribution ($-T\Delta S$) as estimated from the NMODE calculation of the three complexes are also shown in Table 1. The $-T\Delta S$ values for C1 (33.46 kcal/mol), C2 (56.81 kcal/mol), and C3 (48.84 kcal/mol) complexes suggest that the conformational change plays an important role to the different mode of binding of CD44-HA$_{16}$ interaction. Among the three complexes entropic contribution in noted to be more for C2 complex, which might be due to the higher flexibility of the C-terminal residues of CD44 at this complex, whereas the least entropic contribution of C1 is due to its conformational rigidity as evident from the RMSD plot (see Fig. 4). Looking at the contributions of hydrophobic/apolar(ΔG_{vdw} + ΔG_{SA}) and polar/electrostatic (ΔG_{EEL} + ΔG_{PB}) interactions to the total free energy of binding of the three complexes, the magnitude of the apolar contribution (ΔG_{vdw} + ΔG_{SA}) is noted to be dominated. Due to the unfavorable electrostatic interaction energy difference between the free form of CD44 and HA$_{16}$ with respect to its complexed form (ΔG_{EEL}), the contribution of the polar part becomes unfavorable for the C2 and C3 complexes. However, this contribution is noted to be favorable for C1. This infers that among the three binding modes, the parallel and upright modes of binding between CD44 and HA$_{16}$ are dominated by apolar interactions whereas the crystallographic binding mode is favored by both apolar and polar contribution.

3.3. Residue wise decomposition of effective energies

Considering the strong binding affinity of the three binding modes of CD44-HA$_{16}$ complex, in this section we have studied the effective energy per-residue of CD44 by using free energy decomposition method of MM/PBSA approach. As reported by Banerji *et al.*[5], a shallow groove on the CD44 surface constructed by 13 residues namely; Arg22, Try23, Cys58, Arg59, Tyr60, Ile69, Asn75, Ile77, Cys78, Ala79, Ala80, Asn82, and Tyr86 makes close contact with the residues of HA. These residues are hereby named as hyaluronan bind-

ing residues (HBR). Considering that these residues could contribute more towards HA binding, in Table 3 we have listed the decomposed binding free energy contribution of each HBR in the three binding modes of CD44-HA complex. It is evident from the table that all the 13 residues in the three complexes interact favorably with HA$_{16}$. Particularly, for C1, Cys58, Arg59, Tyr60, Ile77, Cys78, Ala79, and Ala80 are noted to interact strongly with HA$_{16}$ as compared to the other HBR. This is in correlation with the experimental evidence[5]. We find from the table that among 13 HBR, Arg22, Ile23, and Arg59 interact most favorably with HA$_{16}$ while forming C2 complex. In addition to this, our calculation further reveals that, apart from these residues the effective contribution arises from Leu88 (–2.85 kcal/mol), Thr89 (–3.22 kcal/mol), Ser90 (–1.63 kcal/mol), Asn91 (–4.28 kcal/mol), Thr92 (–6.32 kcal/mol), Ser93 (–4.35 kcal/mol) and Arg135 (–3.58 kcal/mol). Hence these residues could be considered as a part of the HA binding residues, specifically for the parallel mode. Looking at the effective binding energy of each HBR of C3 complex, we find from Table 3 that out of 13 HBR, Arg22, Tyr23, Arg59, Tyr60, Ile77, and Cys78 have effective contributions towards HA$_{16}$ binding. Additionally, residues like, Asn20 (–2.97 kcal/mol), Ser24 (–1.49 kcal/mol), and Ile124 (–2.03 kcal/mol) interact strongly with HA$_{16}$. It is important to note from the table that Arg22 (actual R41), always contributes

Table 3. Total effective interaction energy (in kcal/mol) of the binding residues of CD44 in the three complexes as obtained from the decomposition method of MM/PBSA calculations

Residue No.	C1	C2	C3
Arg22	–4.34	–3.81	–5.59
Tyr23	–1.59	3.00	–5.46
Cys58	–2.65	–0.06	0.37
Arg59	–2.49	–2.11	–6.63
Tyr60	–5.19	–0.59	–1.80
Ile69	–0.49	–0.07	–0.41
Asn75	1.03	0.03	–0.06
Ile77	–4.14	–0.03	–2.45
Cys78	–3.56	–0.09	–1.43
Ala79	–3.71	–0.11	–0.77
Ala80	–3.67	–0.09	–0.18
Asn82	–0.44	–0.05	–0.01
Tyr86	–0.23	–0.04	–0.08

857

Table 4. Decomposition of effective energy (in kcal/mol) of the binding residues of CD44 in the three complexes as obtained from the MM/PBSA calculations

Residue No.	C1		C2		C3	
	ΔG_{vdw}	$\Delta G_{EEL} + \Delta G_{PB}$	ΔG_{vdw}	$\Delta G_{EEL} + \Delta G_{PB}$	ΔG_{vdw}	$\Delta G_{EEL} + \Delta G_{PB}$
Arg22	−0.29	−4.05	−4.59	0.78	−2.65	−2.93
Tyr23	−0.48	−1.11	−2.99	−0.01	−3.45	−2.01
Cys58	−1.92	−0.73	−0.07	0.01	−0.49	0.86
Arg59	−1.83	−0.65	−0.92	−1.19	−0.40	−6.23
Tyr60	−1.11	−4.07	−1.28	0.70	−1.71	−0.10
Ile69	−0.52	0.03	−0.07	0.00	−0.41	0.00
Asn75	−0.59	1.62	−0.01	0.05	−0.08	0.02
Ile77	−4.03	−0.11	−0.07	0.04	−1.33	−1.12
Cys78	−2.55	−1.02	−0.05	−0.04	−1.12	−0.31
Ala79	−0.67	−3.04	−0.07	−0.05	−0.46	−0.32
Ala80	−1.40	−2.27	−0.06	−0.03	−0.26	0.08
Asn82	−0.48	0.05	−0.02	−0.03	−0.03	0.02
Tyr86	−1.02	0.79	−0.20	0.17	−0.06	−0.02

more to the binding process irrespective to the mode of binding of CD44 and HA_{16}.

The additional insight into the contributions of different interactions to the residue wise effective binding free energy can be further obtained from Table 4. The table reports that the van der Waals interactions are always favorable. Further, the contribution of polar/electrostatic (ΔG_{EEL} + ΔG_{PB}) interactions arising from each binding residue is mostly noted to be favorable in nature. In addition to the estimation of energy contribution arising from the residues of CD44, we have further calculated the contributions arise from the residues of HA_{16}. The residues of HA_{16} that contribute more towards the favorable binding are listed in Table 5. It may be noted from the Table 5 that the residues which contribute more

towards binding fluctuate less as evident from Fig. 5(b). This indicates that the differential binding modes of CD44-HA_{16} is associated with the reduced flexibility of the binding residues of HA_{16}.

3.4. CD44-HA$_{16}$ hydrogen bonds

The favorable polar/electrostatic interactions could assist CD44 and HA_{16} to bound by direct hydrogen bonds. In Fig. 6 we have shown a representative snapshot of CD44-HA_{16} complex in which the residues of CD44 and HA_{16} are bound by direct hydrogen bonds. Further, to get microscopic view of hydrogen bonded interactions between CD44 and HA_{16} in the three complexes, we have estimated the number of hydrogen bonds that formed between the residues of

Table 5. Total effective interaction energy (in kcal/mol) of the residues of HA$_{16}$ in the three complexes that contribute more towards binding as obtained from the MM/PBSA calculations

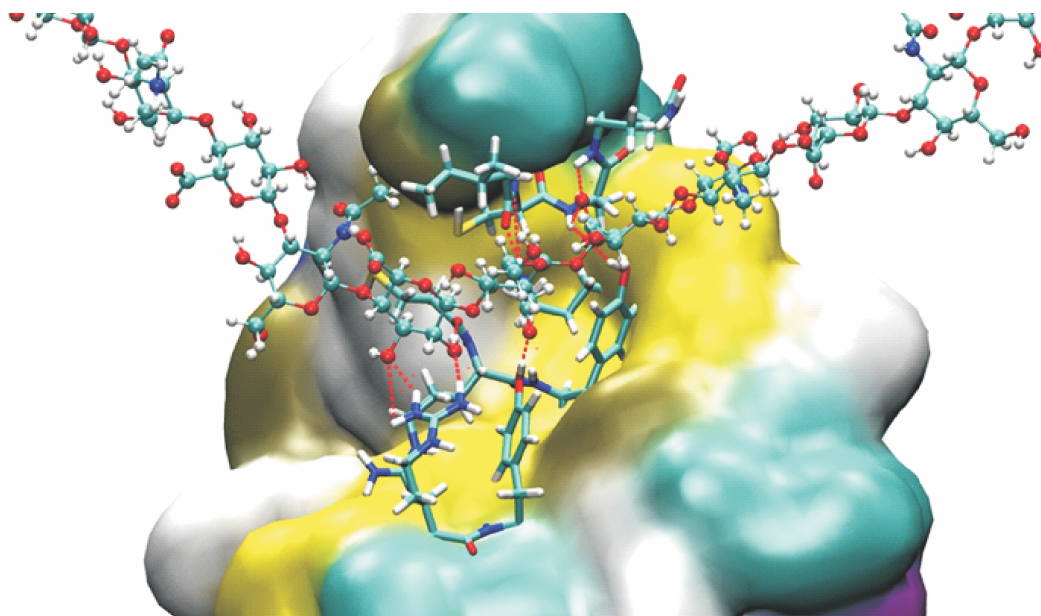| C1		C2		C3	
Residue No.	Total energy	Residue No.	Total energy	Residue No.	Total energy
GlcUA10	−8.43	GlcUA6	−1.87	GlcNAc2	−3.59
GlcNAc11	−4.84	GlcNAc7	−6.16	GlcUA3	−3.53
GlcUA12	−1.59	GlcUA8	−2.46	GlcUA5	−4.11
		GlcUA10	−1.63	GlcNAc6	−1.34
		GlcUA12	−1.29	GlcUA7	−3.95
		GlcNAc15	−2.64	GlcNA8	−2.76
		GlcUA16	−1.65		

Fig. 6. Representative snapshot of the direct hydrogen bonds formed between CD44 and HA$_{16}$ in a simulated CD44-HA$_{16}$ complex. The residues of protein are shown in licorice whereas that of HA in CPK model.

CD44 and HA$_{16}$ over the entire production trajectories. A maximum cut-off distance of 3.5 Å and a minimum Donor-H-Acceptor (D-H-A) angle of 135° are used as geometric criteria to define hydrogen bonds. In Fig. 7(a) we display the distribution of the total number of hydrogen bonds that were formed between CD44 and HA$_{16}$. It is evident from the figure that a significant number of hydrogen bonds is generally formed between CD44 and HA$_{16}$ moiety. It is apparent from the figure that in C1 complex the direct hydrogen bonds formed between CD44 and HA$_{16}$ is more in number than in C2 or C3. To gain further insight on the contribution of CD44/HA as an acceptor or donor, we have separately calculated the number of hydrogen bonds that forms when CD44 acts as an acceptor and HA$_{16}$ as a donor or vice versa. The calculations were carried out over the entire equilibrated trajectories. The average number of such hydrogen bonds as well as the corresponding distribution curves are shown in Table 6 and Fig. 7(b-c) respectively. It is apparent from the figure as well as from the table that in C2 complex, CD44 acts as a good hydrogen bond acceptor (HA$_{16}$ as a good donor) relative to that when it forms C1/C3 complex. This is reflected from the average acceptor-CD44–donor-HA$_{16}$ hydrogen bond number as shown in Table 6. In contrast to this, the relative trend of donor-CD44–acceptor-HA$_{16}$ number follows the order C1 > C3 > C2 as evident from the Fig. 7(c) and the Table 6. It may be additionally noted from Fig. 7(c) that although more number of donor-CD44–acceptor-HA$_{16}$ hydrogen bonds are formed in C1 and C3 complexes, the relatively less number of hydrogen bonds of this kind can form in a much higher extent in C2 and thus the contribution of hydrogen bonds in stabilizing C2 complex cannot be ruled out. Our all such results infer that the formation of direct hydrogen bonds between CD44 and HA$_{16}$ play an important role on the overall binding process.

Table 6. The average number of different types of hydrogen bonds formed between CD44 and HA$_{16}$ as obtained from the equilibrated trajectories of the three complexes

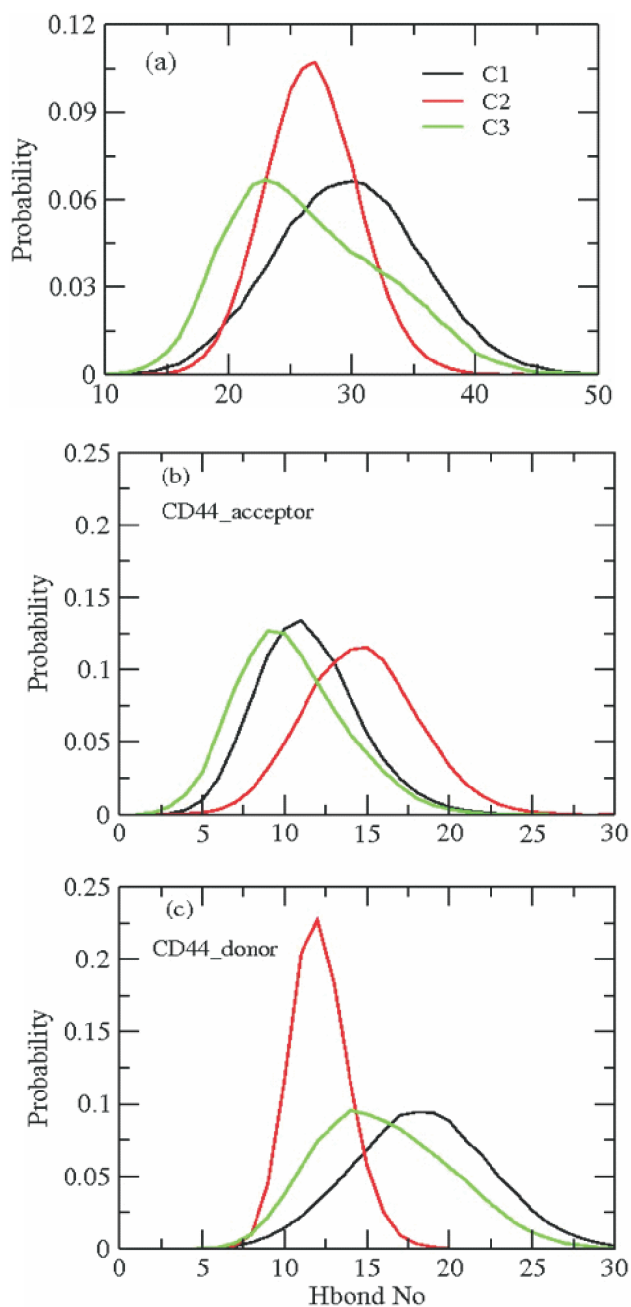| System	H Bond number	
	CD44 acceptor/HA donor	CD44 donor/HA acceptor
C1	11	18
C2	14	12
C3	10	16

Fig. 7. (a) The distribution of the total number of hydrogen bonds formed by CD44 and HA$_{16}$ in the three complexes as calculated from the entire equilibrated trajectories of the three simulations. The distribution plots of (a) the donor CD44-acceptor HA$_{16}$ and (b) the vice versa are also shown.

4. Conclusions

This work was aimed to explore the conformational flexibilities, comparative binding affinities, and the interactions responsible for the different binding modes of CD44-HA$_{16}$ complex namely; crystallographic, parallel and upright mode by using MD simulation and MMPB(GB)SA techniques. Our calculations showed that the conformational flexibility of the complexes plays an important role in the different binding modes of CD44–HA$_{16}$ complex. The CD44–HA$_{16}$ recognition process is found to be governed by the reduced flexibilities of the residues of HA$_{16}$. Our calculations showed that the three binding modes of the complex are favorable and the complexation process is generally governed by the van der Waals interactions. We find from the MM/PBSA calculations that among the three binding modes, although the most favorable one is the parallel mode where the apolar interactions contribute more to the overall binding affinity, existence of strong polar interactions and the direct hydrogen bonding network between CD44 and HA$_{16}$ in their crystallographic mode make its binding second favorable among the three. However, considering the magnitude of difference between the binding affinities of the three modes, all the modes can coexist. From the detailed investigation on the contribution of per residue interactions on the overall free energy of binding, we have identified the residues of HA$_{16}$ and CD44 that contribute most in the three binding modes. Our study reveals that the hyaluronan binding domain residues of link module interact strongly in crystallographic mode, whereas in the other two modes their contribution is relatively less. Importantly, our study showed that arg22 (originally R41) plays significant role in stabilizing all the binding modes of the complex.

References

1. B. P. Toole, *Nat. Rev. Cancer*, 2004, **4(7)**, 528.
2. L. Y. Bourguignon, M. Shiina and J. J. Li, *Adv. Cancer Res.*, 2014, **123**, 255.
3. J. M. Louderbough and J. A. Schroeder, *Mol. Cancer Res.*, 2011, **9(12)**, 1573.

4. P. Teriete, S. Banerji, M. Noble, C. D. Blundell, A. J. Wright, A. R. Pickford, E. Lowe, D. J. Mahoney, M. I. Tammy, J. D. Kahmann, L. D. Campbell, A. J. Day and D. G. Jackson, *Molecular Cell*, 2004, **13(4)**, 483.

5. S. Banerji, A. J. Wright, M. Noble, D. J. Mahoney, I. D. Campbell, A. J. Day and D. G. Jackson, *Nat. Struct. Mol. Biol.*, 2007, **14(3)**, 234.

6. S. Ogino, N. Nishida, R. Umemoto, M. Suzuki, M. Takeda, H. Terasawa, J. Kitayama, M. Matsumoto, H. Hayasaka, M. Miyasaka and I. Shimada, *Structure*, 2010, **18(5)**, 649.

7. M. Takeda, S. Ogino, R. Umemoto, M. Sakakura, M. Kajiwara, K. N. Sagahara, H. Hayasaka, M. Miyasaka, H. Terasawa and I. Shimada, *J. Biol. Chem.*, 2006, **281**, 40089.

8. A. J. Favreau, C. E. Faller and O. Guvench, *Biophys. J.*, 2013, **105(5)**, 1217.

9. J. Vuorio, I. Vattulainen and H. Martinez-Seara, *PLOS*, 2017, **13(7)**, e1005663.

10. D. A. Case, T. A. darden, T. E. Cheatham, C. L. Simmerling III, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang and K. M. Merz, AMBER 12, University of California, San Francisco, CA, USA, 2012.

11. P. H. Hünenberger, *Advance Computer Simulation*, 2005, **173**, 105.

12. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.

13. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23(3)**, 327.

14. U. Essmann, L. Perera and M. L. Berkowitz, *J. Chem. Phys.*, 1995, **103**, 8577.

15. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theor. Comput.*, 2015, **11(8)**, 3696.

16. K. N. Kirschner, A. B. Yongya, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, **29**, 622.

17. W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926.

18. D. W. Wright, S. Wan, C. Meyer, H. V. Vlijmen, G. Tresadern and P. V. Coveney, *Sci. Rep.*, 2019, **9**, 6017.

19. S. Genheden and U. Ryde, *Expert Opin. Drug Discov.*, 2015, **10**, 449.

20. C. Wang, P. H. Nguyen, K. Pham, D. Huynh, T. N. Le, H. Wang, P. Ren and R. Luo, *J. Comput. Chem.*, 2016, **37**, 2436.

21. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33(12)**, 889.

22. A. Onufriev, D. Bashford and D. A. Cash, *Proteins*, 2004, **55(2)**, 383.

23. J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20(2)**, 217.

24. M. F. Sanner, A. J. Olson and J. C. Spehner, *Biopolymers*, 1996, **38(3)**, 305.

25. D. A. Case, "Molecular dynamics and normal mode analysis of biomolecular rigidity", Rigidity Theo. App. Fundamental Material Research, Springer, Boston MA, 2002, 329-344.

26. B. R. Miller III, T. D. McGee (Jr.), J. M. Swalis, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theo. Comput.*, 2012, **8(9)**, 3314.

27. T. Hou, W. Zhang, D. A. Case and W. Wang, *J. Mol. Biol.*, 2008, **376**, 1201.

28. H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330(4)**, 891.