

---

# 14 DNA–Drug Interactions: A Theoretical Perspective

*B. Jayaram, Tanya Singh, and Marcia O. Fenley*

## CONTENTS

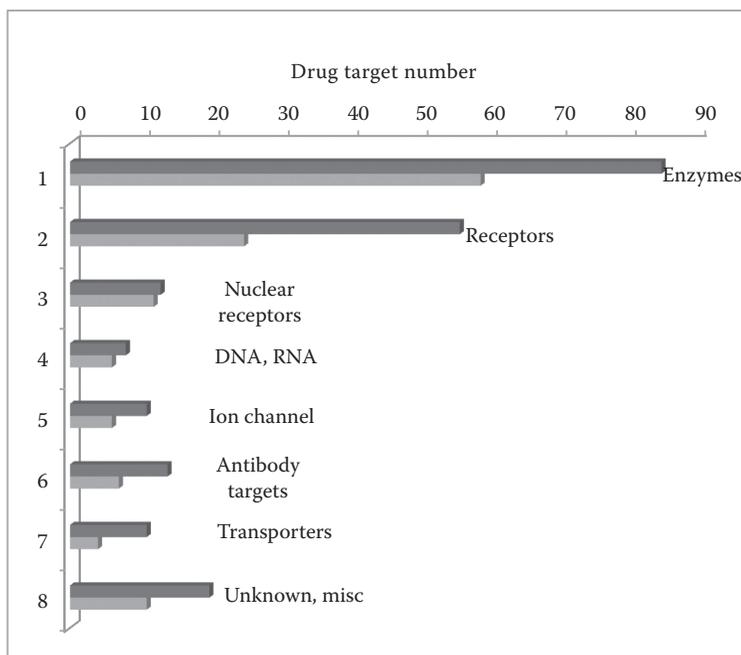
14.1 Introduction .....	317
14.1.1 Drug Targets from Genomic Information .....	318
14.1.2 Optimum Size of DNA as a Target and the Drugs for Specific Binding .....	318
14.2 Background and Technique .....	320
14.2.1 DNA Structure-Based Drug Discovery .....	320
14.2.2 Energetics of DNA–Drug Binding .....	321
14.2.3 Web Tools for Modeling DNA and Drug–DNA Complexes .....	327
14.3 Perspectives and Conclusion.....	327
Acknowledgments.....	327
14.4 Appendix .....	328
References.....	332

The increasing availability of genomic information coupled with the advances in predictive computational tools for characterizing structure, electrostatics, and dynamics of DNA and for estimating DNA–ligand binding free energies usher in an era of DNA-targeted computer-aided drug discovery to combat infectious diseases and noncommunicable disorders, with higher levels of reliability.

## 14.1 INTRODUCTION

DNA as a drug target—be it the DNA of humans, as in cancers, or that of infectious agents—proves attractive due to the availability of the well-studied three-dimensional (3D) DNA structures and the predictability of their accessible chemical functional groups. However, the number of known DNA-based drug targets is still very limited in comparison to the protein-based drug targets (Figure 14.1). The number of available structures of DNA–drug complexes is also small relative to protein–drug complexes deposited in the RCSB database [1] (as shown in Figure 14.1 and Table 14.1), which indicates a heavy underrepresentation of DNA in the structural databases. Two concurrent developments, viz., increasing the availability of genomic sequences and advances in drug delivery systems, are expected to change this scenario drastically in the coming years.

Q1



**FIGURE 14.1** Different therapeutic classes of drug targets for drugs in the market [1–4]. The red bar indicates the total number of known targets for that class and the green bar indicates the number of three-dimensional structures available in the RCSB database [1] for that class. The targets correspond to approved drugs.

### 14.1.1 DRUG TARGETS FROM GENOMIC INFORMATION

Figure 14.2a and b helps explain how the knowledge of genomic sequences can aid in identifying drug targets. For example, based on the knowledge of the genomic information of the malarial parasite *P. falciparum*, three DNA sequences, viz., (a) TGCATGCA, (b) GTGTGCACAC, and (c) GCACGCGTGC, have been identified as regulatory and essential for the functioning of the organism [6,7]. Naturally, if drugs are to be developed against these sequences, one wishes to know the frequency with which these sequences occur in humans in order to ensure that the drugs do not hamper the normal functioning of humans. The availability of genomic sequences, preferably with annotation, makes these choices feasible.

### 14.1.2 OPTIMUM SIZE OF DNA AS A TARGET AND THE DRUGS FOR SPECIFIC BINDING

The genomic sequences also bring to light other considerations. Figure 14.2a suggests that drugs have to bind to at least 16-mers to 18-mers or longer DNA sequences in order to ensure selectivity. Examination of the genomic sequences in humans (Figure 14.2b) indicates that the drugs have to cover at least 18 base pairs (bp) in order to uniquely bind to their targets with both high affinity and specificity.

**TABLE 14.1**  
**Number of Three-Dimensional Structures of DNA and DNA–Drug Complexes Reported in the Nucleic Acid Database (NDB) [5]**

Number	Type of Complex/Binding	Total Number of PDB and NDB Entries
1	DNA	2689
2	DNA minor groove binders	109
3	DNA major groove binders	11
4	DNA intercalators	148
5	DNA–protein complexes	1692
6	Protein–ligand complexes	22,312

*Note:* The protein–ligand statistics was taken from the RCSB database [1].

These numbers are reminiscent of the classical genetic switch [8] of the  $\lambda$  cro and  $\lambda$  repressor-operator systems wherein the operators are 17 bp long. Similarly, although each zinc-finger DNA-binding motif covers only 3 bp, multiple zinc-finger motifs bind in tandem to cover longer DNA sequences for transcriptional activation/regulation [9–15].

The above discussion brings to focus the optimal dimensions a drug molecule must possess for specific binding to its DNA target. The axial distance of 20 bp (approximately two turns) of B-DNA is  $\sim 68$  Å. The contour length along the grooves is longer ( $\sim 90$  Å). If the designed drug is a groove binder, its end-to-end distance has to be  $> 90$  Å. Typical lengths of some known DNA-binding ligands (drugs) are shown in Table 14.2 along with the target DNA base sequences. This data suggests the need to design drugs that are at least four times longer than netropsin, with overall molecular weights in the range of 1600–1800 Da, to ensure specific binding to unique DNA targets (Figure 14.3).

The length considerations of DNA-targeted drugs bring to fore nonconformity with Lipinski's rules [21,22] and drug delivery issues. A possible unexplored solution that has considerable promise is the design of monomeric drugs that bind to shorter sequences but could then form homo- or heterooligomeric drug(s) when binding to its DNA target. We propose that, for instance, four netropsin-like molecules with different sequence specificities and with an end-on coupling can easily cover any unique target sequence akin to the well-known binding of proteins to DNA as either dimers [8] or higher-order oligomers [23]. Recent advances in nano-biotechnology are expected to overcome hurdles that are still present in the delivery of drug molecules to their target site. These possibilities strongly affirm the potential for the success of novel drugs targeted to DNA.

Other design considerations, apart from length/size, include the chemical information content in the DNA grooves. The minor groove has a lower informational content than the major groove in the sense that both AT and TA base pairs project electronegative atoms into the minor groove. There exists a subtle polarity of the electrostatic potentials between GC and CG sequences in the minor grooves. Most proteins

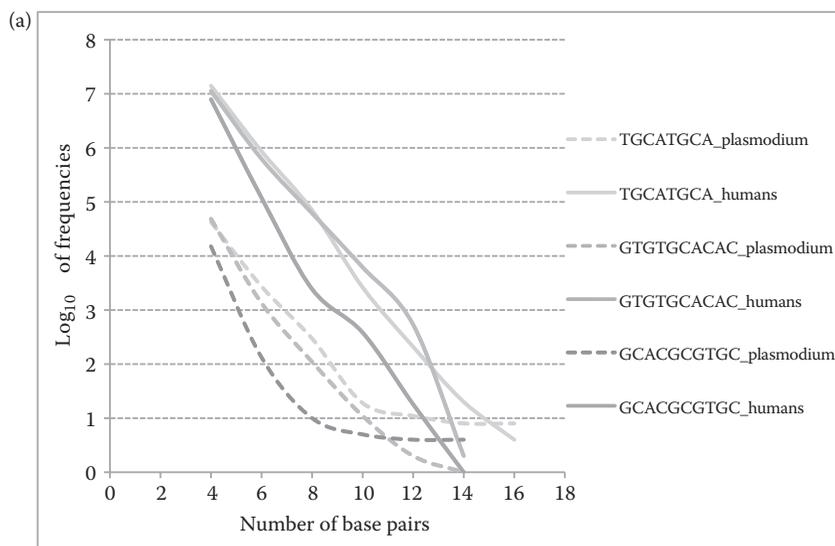
Q2

target the major groove. However, most known drugs target the minor groove enabling a snug fit between the binding partners. The wider major groove requires bulkier drugs with larger molecular weights. Cationic drugs have been the popular choice for binding to AT-rich minor groove regions while intercalators seem to prefer GC steps [24] (see, however, Ref. 25). One obvious choice apart from Dervan's architectures [26–29] is to bind to the minor groove in AT-rich regions and intercalate into GC regions. A rational combination of this strategy to cover the desired DNA sequences of any length is yet to emerge.

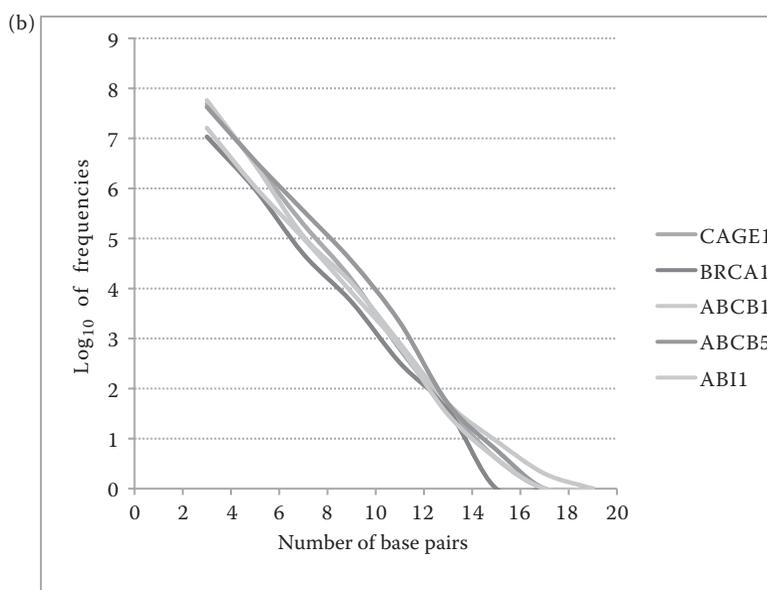
## 14.2 BACKGROUND AND TECHNIQUE

### 14.2.1 DNA STRUCTURE-BASED DRUG DISCOVERY

In the absence of x-ray crystallographic or NMR structures of target DNA sequences, drug design efforts require sophisticated molecular simulation techniques to capture the subtle base sequence effects on the groove widths and the spatial disposition of the chemical functional groups. Also, very accurate computational tools for determining the



**FIGURE 14.2** (a) Logarithm of the frequencies of the occurrence of base sequences of lengths 4–18 base pairs in *Plasmodium falciparum* and in humans embedding a regulatory sequence TGCATGCA (shown in green), GTGTGCACAC (blue), and GCACGCGTGC (orange) or parts thereof, of the plasmodium. The solid lines and the dashed lines correspond to humans and plasmodium, respectively. Curves lying between 0 and 1 on the log scale indicate occurrences in single digits. (b) Logarithm of the frequencies of occurrence of base sequences from 3 to 18 base pairs in humans embedding a regulatory sequence AAGCTGTCATTA or parts thereof of a cancer causing CAGE1 gene [16], GACTGAGTCAA or parts thereof of a cancer causing BRCA1 gene [17], CTCTAAGTCAT or parts thereof of a cancer causing gene ABCB1 [18], GATATGTTAAAGC or parts thereof of a cancer causing gene ABCB5 [19], and CTTCTGGGAA or parts thereof of a cancer-causing gene ABII [20].



**FIGURE 14.2** (Continued)

electrostatic potentials of DNA with different base sequences are necessary for reliable drug design efforts. Fortunately, these have become available in the recent years.

The Ascona B-DNA consortium [30,31] was formed with the objective of developing state-of-the-art molecular dynamics simulation protocols to describe DNA accurately in aqueous media and the simulation results vis-à-vis 3D structures in the PDB/NDB databases are very encouraging. Statistical mechanical theory that can utilize the molecular simulation trajectories to determine the energetics of binding has also been worked out (see Appendix). Empirical potential functions to analyze single structures/snapshots of molecular dynamics or Monte Carlo simulations that yield results in good correlation with pertinent experimental data have also been developed [32–38].

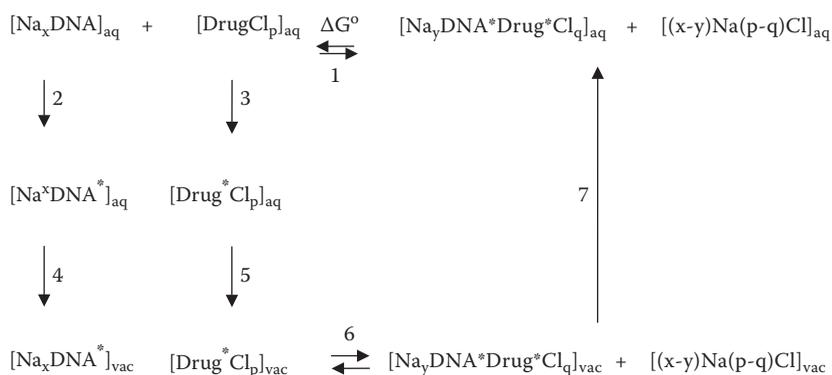
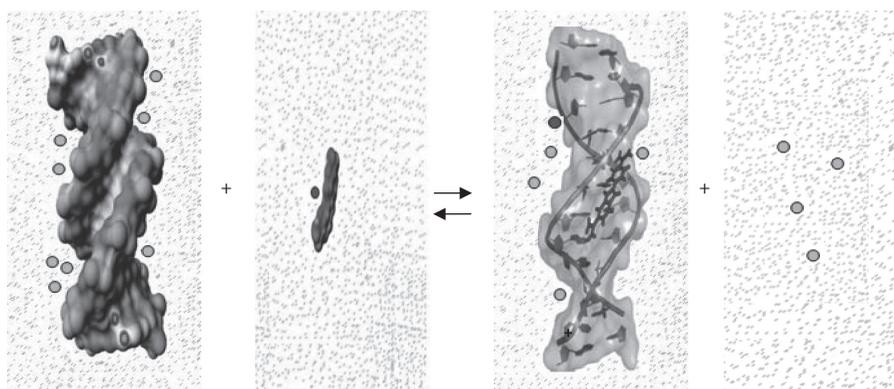
Advances in finite difference Poisson–Boltzmann (FDPB) methodology [39–44] and the analyses of electrostatic potentials of DNA [45–48] obtained as solutions to the nonlinear PB equation [39] now make it possible to elucidate potential recognition/binding sites that result from the base sequence and/or from the phosphodiester backbone. The computed surface electrostatic potential profiles of DNA and its cationic drug partner allow one to assess the electrostatic potential complementarities from a qualitative perspective. However, the development of more quantitative electrostatic potential metrics is vital for drug design.

### 14.2.2 ENERGETICS OF DNA–DRUG BINDING

A drug molecule that competes with a regulatory protein has to generate a binding free energy in the range of  $-9$  to  $-15$  kcal/mol [49]. Minor groove binders have to achieve

**TABLE 14.2**  
**Typical Dimensions of Some Known DNA-Binding Drugs**

Number	PDBID	Drug Name	Action	Molecular Formula	End-to-End Dimensions (in Å)	Name of Atoms	Base Pairs Spanned
1	121D	Netropsin	Antitumor, antiviral	$C_{18}H_{26}N_{10}O_3$	18.9	N1-N10	AATTT
2	3FT6	Proflavine	Anti-infectives	$C_{13}H_{11}N_3$	9.6	N15-N16	CG
3	1AU5	Cisplatin	Anticancer antibiotic	$H_6Cl_2N_2Pt$	2.9	N1-N2	GG
4	227D	Guanyl bisfuramide	Active against <i>P. carinii</i>	$C_{18}H_{18}N_4O$	12.8	N1-N1'	AATT
5	1D63	Berenil	Antitrypanosomal	$C_{14}H_{15}N_7$	12.3	NA-NA'	AATT
6	182D	Nogalamycin	Antitumor	$C_{39}H_{49}NO_{16}$	12.7	O1'-C32	TG
7	202D	Menogaril	Antitumor topoisomerase II poison	$C_{28}H_{31}N_{10}$	13.07	C5M-C7M	GT
8	408D	Imidazole-pyrrole polyamide	—	$C_{31}H_{42}N_{11}O_5$	21.50	N7-C26	AGTACT



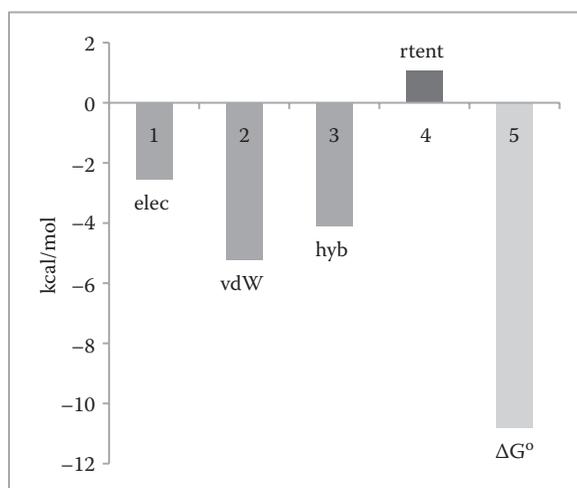
**FIGURE 14.3** A thermodynamic cycle to study the energetics of DNA–drug binding. The formation of the DNA–DAPI complex (PDB id: 1D30), a typical minor-groove-binding drug–DNA complex is shown. The electrostatic potential of the face of the drug that contacts the minor groove is very positive (blue and green) as opposed to the remaining surface, which is mostly neutral. Note that the most negative (red and yellow) electrostatic potential of DNA for this AT-rich sequence is in the deep and narrow minor groove rather than along the sugar–phosphate backbone. This strong negative potential originates from both the anionic sugar–phosphate backbone and electronegative O2 of thymine and N3 atoms of adenine. The binding position of the drug is shown. The curved surface of the drug (in blue sticks) fits tightly in the minor groove of the DNA. The color scheme used in these surface electrostatic potential maps is as follows: yellow is the most negative and green is the most positive. White is neutral. Red and blue represent negative and positive electrostatic potentials, respectively.

this via a snug fit and electrostatic complementarity without much structural distortion and intercalators via stacking in order to overcome the energy penalty associated with local unwinding/structural adaptation of the DNA. In order to design drugs that target DNA with enhanced binding specificity and affinity, many groups have proposed that a better understanding of different noncovalent interactions, such as van der Waals, and electrostatics along with structural information is warranted. An energy component

analysis of the binding of several known DNA-binding drugs with an empirical potential function is shown in Figure 14.4. Except for the entropies, all the energy components, electrostatics, van der Waals, and hydrophobic/cavity terms, seem to favor binding to varying degrees. Design efforts have long been focused on optimizing the number of hydrogen bonds and good steric fits in the grooves.

Since DNA and the cationic organic drugs that either intercalate into the DNA or bind to its grooves are charged molecules, it is not surprising that nonspecific and long-range electrostatic interactions have been found to be vital to the binding process. It is conceivable that in the initial binding step, the long-range and nonspecific electrostatic interactions help steer these charged molecules into their initial or transient encounter complex. At this stage of the binding process, the fine atomic details of the charge distribution do not matter and electrostatics is favorable and dominated by the Coulomb attraction between the oppositely charged binding partners [50]. At a later stage of binding that leads to the final docked bound state, short-range interactions such as van der Waals/hydrophobic contributions appear to play a large role in driving the formation of the stable complex (Figure 14.4). Also, short-range and directional electrostatic interactions such as hydrogen bonding and salt bridges play a critical role for the specificity of the final complex formation.

No experimental approach is available to directly measure or quantify the contribution of electrostatic interactions to the association reaction between biomolecules, at least with the present biophysical techniques (however, see Ref. 51). Thus, different methods have been proposed to infer information about the role of electrostatics in the



**Q6** **FIGURE 14.4** An energy component analysis of DNA–drug binding through PreDDICTA [32, 33], an empirical scoring function, for minor groove binders. The preDDicta method refers to the final state analysis in step 1 of Figure 14.3. The figure depicts a consensus view of favorable (blue) and unfavorable (red) energy components contributing to the binding free energy (green) emerging from several systems (PDB ids: 127d, 264d, 109d, 1d63, 2dbe, 121, 2dnd, 227d, 298d, 289d, 1fmq, 1eel, 1fms, 1pp). Elec, electrostatics; vdW, van der Waals; hyb, hydrophobic; rtent, rotational translational entropy. Binding free energy estimate (BFEE) =  $\Delta G^\circ$ .

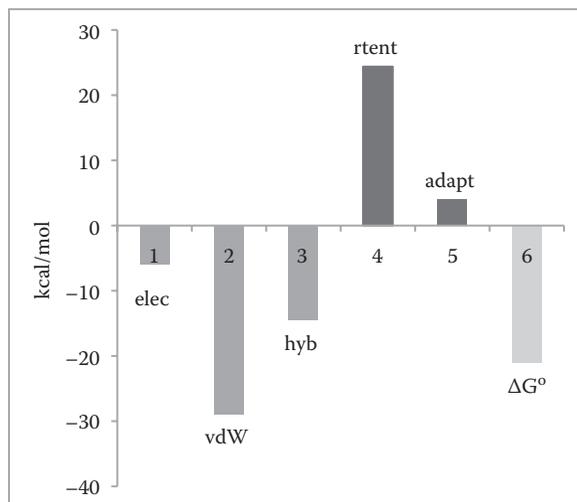
binding process. Their quantitative and qualitative aspects, merits, and limitations are briefly discussed below.

A very well-established approach that is widely used to extract the electrostatic contribution to the biomolecular binding processes relies on measuring the equilibrium binding constant at varying added salt concentrations. The classic log–log plot of the observed binding constant ( $K_{\text{obs}}$ ) as a function of the 1:1 salt concentration  $[M^+]$  often portrays a linear relationship between these two quantities. This is true at least over a moderate range of 1:1 salt concentration and in the absence of competing multivalent ions for charged biomolecular complexes [52]. The slope of this linear log–log plot is referred to as  $SK_{\text{obs}}$  ( $= \text{dlog } K_{\text{obs}}/\text{dlog } [M^+]$ ) in the literature. This linear log–log plot has become a signature of the polyelectrolyte effect for charged ligand–nucleic acid complexes [53–56]. The electrostatic contribution to the binding free energy is then obtained as  $\Delta G_{\text{el}} = SK_{\text{obs}} \ln [M^+]$  [54,56,57], as inferred from counterion condensation theory (CCT). The  $SK_{\text{obs}}$  metric is often considered to reflect the net charge on the cationic drug and the number of “condensed” counterions released from DNA upon drug binding, which is entropically favorable. A very strong support for the use of this simple electrostatic model, where the fine atomic details of the binding molecules and dielectric discontinuity effects are ignored, comes from the fact that CCT can correctly predict the experimentally observed  $SK_{\text{obs}}$  for small charged ligands binding to polymeric DNA or RNA [54,56].

Interestingly, theoretical studies using the nonlinear Poisson–Boltzmann equation with a formal charge distribution and no dielectric discontinuity show that  $SK_{\text{obs}}$  equals the net charge of the drug in agreement with the CCT prediction [44,50,58–66]. It is important to stress that only the nonlinear solution of the PBE provides  $SK_{\text{obs}}$  values that agree with both CCT and experimental binding data [44,67–72]. However, the linear relationship  $\Delta G_{\text{el}} = SK \ln [M^+]$ , which is now widely used to parse the total binding free energy into electrostatic and nonelectrostatic contributions [57,73–75], does not hold when dielectric discontinuity effects are considered. The importance of field discontinuities in correctly portraying electrostatic features of biomolecules has been shown by several PB studies [41,76]. Moreover, according to this popular relationship between  $SK_{\text{obs}}$  and  $\Delta G_{\text{el}}$ , any drug having the same net charge but different shape and charge distribution will have the same  $\Delta G_{\text{el}}$ . The same is not true according to the all-atom PB approach. These results highlight some issues yet to be resolved in deciphering the electrostatic contribution to the overall binding energetics from continuum or implicit solvent models. This notwithstanding, theory (CCT or PB) provides excellent predictions of  $SK_{\text{obs}}$ . This finding confirms the dominant role of long-range nonspecific electrostatics in controlling the salt sensitivity of the binding energetics. Thus, it is paradoxical and even counterintuitive that a large  $SK_{\text{obs}}$  (in terms of its magnitude) does not necessarily imply a large electrostatic contribution to binding as often assumed in the literature. In fact, many examples in the literature and from our work clearly show the opposite trend. For instance, some nucleic acid-binding proteins containing a large number of anionic residues, such as the tRNA synthetases and elongation factor Tu when bound to its nucleic acid-binding partner, have  $SK_{\text{obs}}$  absolute values that are very small or close to zero, but the electrostatic contribution to binding energy is quite large and unfavorable [41,77,78]. A strong correlation between electrostatic complementarity metrics and the electrostatic binding free energy of nucleic acid complexes is yet to be established.

Some of the issues in theoretical studies that are currently under microscopic scrutiny in diverse laboratories include more quantitative measures of electrostatic potentials, sensitivity of the PB predictions to parameters such as radii, the relative permittivities of solvent and solute, the boundary conditions, molecular surface definition, the compensatory nature of Coulomb, and desolvation interactions. A resolution of these issues is expected soon.

Molecular dynamics (MD) simulation studies on drug–DNA complexes provide additional insights into the structure, dynamics, and energetics of binding [79–88]. Results of a *post facto* molecular mechanics/generalized born surface area (MM/GBSA) analysis [36,89–92] of the MD trajectories of the unbound DNA d(CGCAAATTTGCG)<sub>2</sub>, berenil, and the complex of DNA with berenil are shown in Figure 14.5. While the overall picture remains similar to the empirical energy function analyses (Figure 14.4) on drug–DNA complexes and protein–DNA complexes [93,94], MD trajectory analyses take into account structural adaptation of the interacting molecules, solvation/desolvation effects, and the role of explicit ions in binding. While some care is necessary in calling a component as favorable to binding due to the compensatory nature of several components with opposing effects contributing to binding, electrostatic complementarities, good steric fit of the drug in the grooves, as well as hydrophobic components are highlighted as important for berenil–DNA



**FIGURE 14.5** *Post facto* analyses of molecular dynamics trajectories on DNA, drug, and the complex using MM/GBSA theory and the thermodynamic cycle (steps 2 through 7) in Figure 14.3. An 8ns molecular dynamics simulation was performed on the DNA berenil complex (pdb id: 1d63), on free DNA, and on the drug with explicit solvent and counterions [30]. About 100 structures spaced at equal time intervals were culled from each of the three trajectories and averages of each energy component contributing to binding were computed. The net value of each energy component ( $\Delta E$ ) is calculated as  $\Delta E = [E_{\text{Complex}} - E_{(\text{DNA} + \text{Drug})}]$ . Elec, electrostatics; vdW, van der Waals; hyb, hydrophobic; rtent, rotational translational entropy; adapt, adaptation energy. Binding free energy estimate (BFEE) =  $\Delta G^\circ$ .

binding. Rotational, translational entropies, vibrational/configurational entropies (not computed), and adaptation are expected to be unfavorable to drug–DNA binding. The magnitudes of the components provide a good indication of the order of importance of the energy components favorable to binding and prove valuable in drug design efforts.

### 14.2.3 WEB TOOLS FOR MODELING DNA AND DRUG–DNA COMPLEXES

Theoretical methodologies are increasingly being converted into user-friendly software to facilitate drug design endeavors. Some of the web tools for analyzing DNA structure and for assessing DNA–drug binding are given in Tables 14.3 and 14.4.

## 14.3 PERSPECTIVES AND CONCLUSION

Atomic models and molecular modeling and simulation methodologies of drug binding to DNA have matured to a stage where it is conceivable to generate reliable *in silico* suggestions of candidate molecules to bind to any specific base sequence of DNA. The stage is set for DNA-targeted drug discovery.

## ACKNOWLEDGMENTS

B. Jayaram acknowledges the program support to SCFBio from the Department of Biotechnology, Government of India. Marcia O. Fenley acknowledges the support

**TABLE 14.3**  
**Web Tools for DNA Structure Generation and Analysis**

Tool	Website	Description
Curves <sup>95,96</sup>	<a href="http://gbio-pbil.ibcp.fr/Curves_plus/Curves+.html">http://gbio-pbil.ibcp.fr/Curves_plus/Curves+.html</a>	Program to analyze DNA structure
NUPARM <sup>96–102</sup>	<a href="http://nucleix.mbu.iisc.ernet.in/nuparm/nuparm.shtml">http://nucleix.mbu.iisc.ernet.in/nuparm/nuparm.shtml</a>	Program to analyze sequence-dependent variations in nucleic acid (DNA and RNA) double helices
3DNA <sup>103,104</sup>	<a href="http://rutchem.rutgers.edu/~xiangjun/3DNA/">http://rutchem.rutgers.edu/~xiangjun/3DNA/</a>	Program to analyze, rebuild, and visualize three-dimensional nucleic acid structures
NUCCGEN <sup>95</sup>	<a href="http://nucleix.mbu.iisc.ernet.in/nucgen/index.htm">http://nucleix.mbu.iisc.ernet.in/nucgen/index.htm</a>	Program to generate a curved or nonuniform helix
AMBER <sup>105,106</sup>	<a href="http://ambermd.org/">http://ambermd.org/</a>	Program to generate canonical A- and B-duplex geometries of nucleic acids
DNA sequence to structure <sup>107</sup>	<a href="http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp">http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp</a>	Program to generate canonical A and B DNA and molecular dynamics-averaged DNA structure

**TABLE 14.4**  
**DNA–Ligand and DNA–Protein Docking and Scoring Software**

Docking Softwares	Website	Description
Surflex <sup>108</sup>	<a href="http://www.tripos.com/index.php?family=modules,SimplePage,,&amp;page=surflex_dock&amp;s=0">http://www.tripos.com/index.php?family=modules,SimplePage,,&amp;page=surflex_dock&amp;s=0</a>	DNA–ligand docking software
Autodock 4.0 <sup>109</sup>	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>	DNA–ligand docking software
DNA–ligand docking <sup>110</sup>	<a href="http://www.scfbio-iitd.res.in/dock/dnadock.jsp">http://www.scfbio-iitd.res.in/dock/dnadock.jsp</a>	DNA–ligand docking software
GOLD <sup>111–117</sup>	<a href="http://www.ccdc.cam.ac.uk/products/life_sciences/gold/">http://www.ccdc.cam.ac.uk/products/life_sciences/gold/</a>	Software for searches of databases for DNA–binding compounds and for DNA–protein docking
HADDOCK <sup>118</sup>	<a href="http://www.nmr.chem.uu.nl/haddock/">http://www.nmr.chem.uu.nl/haddock/</a>	DNA–protein docking software
Escher NG <sup>119</sup>	<a href="http://users.unimi.it/~ddl/escherng/index.htm">http://users.unimi.it/~ddl/escherng/index.htm</a>	DNA–protein docking software

from NIH-GM078538-01 (PI: Dr. Michael S. Chapman, OHSU, Co-PI: MOF, FSU) and SBIR NIH 2R44GM073391-02 (PI: Dr. Alexander H. Boschitsch).

## 14.4 APPENDIX

A statistical mechanics theory for DNA–drug binding in aqueous salt media on the lines of protein–DNA binding [120] and protein–ligand binding [121] is presented here.

Let D (denoting DNA) and dr (denoting drug) be the reactants and D\*dr\*, the product of binding in aqueous salt medium.

$$[D]_{\text{aq}} + [\text{dr}]_{\text{aq}} = [D^* \text{dr}^*]_{\text{aq}} \quad (14.1)$$

At equilibrium,

$$\mu_{\text{D,aq}} + \mu_{\text{dr,aq}} = \mu_{\text{D}^*\text{dr}^*,\text{aq}} \quad (14.2)$$

where  $\mu_{\text{D,aq}}$  is the chemical potential of species D in the ionic solvent medium (partial molar Gibbs free energy) and  $\mu_{\text{D,aq}}^\circ$  is its standard chemical potential, that is, under condition of 1 bar in gaseous state and 1 M (designated as  $C^\circ$ ) in liquid state.

$$\begin{aligned} \mu_{\text{D,aq}}^\circ + RT \ln(\gamma_{\text{D}} C_{\text{D}}/C^\circ) + \mu_{\text{dr,aq}}^\circ + RT \ln(\gamma_{\text{dr}} C_{\text{dr}}/C^\circ) \\ = \mu_{\text{D}^*\text{dr}^*,\text{aq}}^\circ + RT \ln(\gamma_{\text{D}^*\text{dr}^*} C_{\text{D}^*\text{dr}^*}/C^\circ) \end{aligned} \quad (14.3)$$

where  $\gamma_{\text{D}}$  is the activity coefficient of species D and  $C_{\text{D}}$  is its concentration. The standard molar Gibbs free energy of the reaction (standard absolute molar Gibbs free energy of binding) is

$$\begin{aligned}\Delta G_{\text{aq}}^{\circ} &= \mu_{\text{D}^*\text{dr}^*\text{aq}}^{\circ} - (\mu_{\text{D}\cdot\text{aq}}^{\circ} + \mu_{\text{dr}\cdot\text{aq}}^{\circ}) \\ &= -RT \ln[\gamma_{\text{D}^*\text{dr}^*} C_{\text{D}^*\text{dr}^*} C^{\circ} / (\gamma_{\text{D}} C_{\text{D}})(\gamma_{\text{dr}} C_{\text{dr}})] = -RT \ln K_{\text{aq}}\end{aligned}\quad (14.4)$$

In terms of the canonical partition function ( $Q$ )

$$\begin{aligned}\Delta G_{\text{aq}}^{\circ} &= \Delta A_{\text{aq}}^{\circ} + P\Delta V_{\text{aq}}^{\circ} = -RT \ln K_{\text{aq}} \\ &= -RT \ln \left[ \frac{Q_{\text{D}^*\text{dr}^*\text{aq}}}{(N_{\text{A}} Q_{\text{W}})} \right] / \left[ \frac{Q_{\text{D}\cdot\text{aq}}}{(N_{\text{A}} Q_{\text{W}})} \right] \left[ \frac{Q_{\text{dr}\cdot\text{aq}}}{(N_{\text{A}} Q_{\text{W}})} \right] \Bigg] \\ &\quad + P\Delta V_{\text{aq}}^{\circ}\end{aligned}\quad (14.5)$$

where  $\Delta A^{\circ}$  is the standard Helmholtz free energy of the reaction,  $P\Delta V_{\text{aq}}^{\circ}$  is the pressure volume correction to Helmholtz free energy in the solvent medium,  $K_{\text{eq, aq}}$  is the equilibrium constant for the reaction in (1),  $N_{\text{A}}$  is the Avogadro number, and  $Q_{\text{W}}$  denotes the partition function for pure solvent (water).

Q3

Assuming that translations and rotations are separable from intrasolute degrees of freedom as well as those of the solvent,

$$\begin{aligned}\Delta G_{\text{aq}}^{\circ} &= -RT \ln \left\{ \frac{Q_{\text{D}^*\text{dr}^*}^{\text{rot}} Z_{\text{D}^*\text{dr}^*}^{\text{int}} Q_{\text{D}^*\text{dr}^*}^{\text{el}} N_{\text{A}} Q_{\text{W}}}{(Q_{\text{D}}^{\text{tr}} Q_{\text{D}}^{\text{rot}} Z_{\text{D}\cdot\text{aq}}^{\text{int}} Q_{\text{D}}^{\text{el}})(Q_{\text{dr}}^{\text{tr}} Q_{\text{dr}}^{\text{rot}} Z_{\text{dr}\cdot\text{aq}}^{\text{int}} Q_{\text{dr}}^{\text{el}})} \right\} + P\Delta V_{\text{aq}}^{\circ}\end{aligned}\quad (14.6)$$

The superscript “int” denotes the internal contribution, and  $Z^{\text{int}}$  is the configurational partition function. It includes contributions from the intermolecular interactions and internal motions as well as solvation (hydration) effects. The translational and rotational terms have been separated out.

$Z^{\text{int}}$  is determined via

$$\begin{aligned}Z_{\text{D}\cdot\text{aq}}^{\text{int}} &= \int \cdots \int \exp\{-E(X_{\text{D}}^{\text{N}}, X_{\text{W}}^{\text{M}}) / k_{\text{B}} T\} dX_{\text{D}}^{\text{N}} dX_{\text{W}}^{\text{M}} \\ &= \langle \exp(E(X_{\text{D}}^{\text{N}}, X_{\text{W}}^{\text{M}}) / k_{\text{B}} T) \rangle\end{aligned}\quad (14.7)$$

$X_{\text{D}}^{\text{N}}$  and  $X_{\text{W}}^{\text{M}}$  represent the configurational space accessible to the solute D and solvent W, respectively, in the presence of each other.  $E(X_{\text{D}}^{\text{N}}, X_{\text{W}}^{\text{M}})$  denotes the total potential energy of the system describing nonidealities.

The electronic partition function  $Q^{\text{el}}$  is assumed to be unity for noncovalent associations,

$$Q_{\text{D}}^{\text{el}} = Q_{\text{dr}}^{\text{el}} = Q_{\text{D}^*\text{dr}^*}^{\text{el}} = 1\quad (14.8)$$

The standard free energy can be expressed as a sum of external (translational and rotational) and internal (intramolecular, intermolecular, and solvation) contributions.

$$\begin{aligned}\Delta G^{\circ} &= -RT \ln \left[ \frac{Q_{\text{D}^*\text{dr}^*}^{\text{tr}} N_{\text{A}}}{(Q_{\text{D}}^{\text{tr}} Q_{\text{dr}}^{\text{tr}})} \right] - RT \ln \left[ \frac{Q_{\text{D}^*\text{dr}^*}^{\text{rot}}}{Q_{\text{D}}^{\text{rot}} Q_{\text{dr}}^{\text{rot}}} \right] \\ &\quad - RT \ln \left[ \frac{(Z_{\text{D}^*\text{dr}^*\text{aq}}^{\text{int}} Q_{\text{W}})}{(Z_{\text{D}\cdot\text{aq}}^{\text{int}} Z_{\text{dr}\cdot\text{aq}}^{\text{int}})} \right] + P\Delta V_{\text{aq}}^{\circ}\end{aligned}\quad (14.9)$$

Equation 14.9 is an exact expression for evaluating binding free energies for noncovalent associations in aqueous medium. The first two terms on the right-hand side of Equation 14.9 can be computed analytically. The third term is accessible to free energy

molecular simulations configured in the canonical ensemble such as the perturbation method, thermodynamic integration, potential of mean force method, etc., albeit they are computationally expensive for a single ligand and not practical in a high-throughput sense even on supercomputers.

Here, some simplifications are considered to bring the binding free energy computations into the feasibility domain. The molecular translational partition function of D is

$$q_{\text{D}}^{\text{tr}} = V/\Lambda_{\text{D}}^3 = V/(h^2/2\pi m_{\text{D}} k_{\text{B}} T)^{3/2} \quad (14.10)$$

where  $V$  is the volume,  $\Lambda_{\text{D}}$  is the thermal wavelength of D,  $h$  is the Planck's constant,  $k_{\text{B}}$  is the Boltzmann constant,  $T$  is the temperature, and  $m$  is the mass.

The molar partition function of D is

$$Q_{\text{D}}^{\text{tr}} = (q_{\text{D}}^{\text{tr}})^{N_{\text{A}}} \quad (14.11)$$

The volume  $V$  has been included in the translational part consistent with ideal gas statistical mechanics. This would require that the  $Z^{\text{int}}$  be divided by  $V$  to quantify nonidealities (excess free energies). The translational part of the free energy in Equation 14.9 is now given by the Sackur–Tetrode [10] equivalent as

$$\begin{aligned} \Delta G_{\text{tr}}^{\circ} &= -RT \ln \left[ (N_{\text{A}}/V) (\Lambda_{\text{D}}^3 \Lambda_{\text{dr}}^3 / \Lambda_{\text{D}^*\text{dr}^*}^3) \right] \\ &= -RT \ln \left[ (N_{\text{A}}/V) (h^2/2\pi k_{\text{B}} T)^{3/2} \{ m_{\text{D}^*\text{dr}^*} / (m_{\text{D}} m_{\text{dr}}) \}^{3/2} \right] \end{aligned} \quad (14.12)$$

The expression in square brackets in Equation 14.12 is dimensionless.  $(N_{\text{A}}/V)$  may be replaced by a concentration term ensuring that in the transfer to aqueous medium, standard free energies are recovered with the reference state with a molar concentration of unity. This expression is the same whether in gas phase or liquid phase, provided that the translational and rotational motions of the solute are unaffected by the solvent. This will be true only in a continuum, friction-less solvent influencing the position-dependent potential energy but not the velocity-dependent kinetic energy of the solute. Hence, in a transfer process (an experiment involving transfer of species D from one phase to another such as from gas phase to liquid phase, octanol to water, etc.), this term cancels out. In the binding processes however, no such cancellation occurs. Also, if D, dr, and D\*dr\* could be seen as a collection of nonbonded monoatomic particles, then again the translational partition function for each species could be written as a product of the individual partition functions of the constituent atoms and since the number of atoms is conserved during binding, these terms would cancel out. Again, this is not so for polyatomic species where the mass in translational partition function  $m_{\text{D}} = \sum_i m_i$  is evaluated as a sum of the masses of the constituent atoms. It is thus recommended that Sackur–Tetrode equation be applied not in the aqueous medium directly where it is invalid but upon transfer to vacuum via a suitable thermodynamic cycle shown in Figure 14.3.

Similar arguments apply to the rotational partition functions. Separating the rotational part from internal motions implies working under the rigid rotor approximation.

The molecular rotational partition function of D is

$$q^{\text{rot}} = \sigma^{-1} (k_B T / hc)^{3/2} (\pi / I^a I^b I^c)^{1/2} \quad (14.13)$$

$$\Delta G^{\circ}_{\text{rot}} = -RT \ln \left[ (\sigma_D \sigma_{\text{dr}} / \sigma_{D^* \text{dr}^*}) (1 / (8\pi^2)) (h^2 / 2\pi k_B T)^{3/2} \right. \\ \left. * \left\{ (I^a_{D^* \text{dr}^*} I^b_{D^* \text{dr}^*} I^c_{D^* \text{dr}^*}) / (I^a_D I^b_D I^c_D I^a_{\text{dr}} I^b_{\text{dr}} I^c_{\text{dr}}) \right\}^{1/2} \right] \quad (14.14)$$

$I^a_D$ ,  $I^b_D$ , and  $I^c_D$  are the components of moments of inertia of species D along the principal axes and  $s_D$  its symmetry number.

Considering Equations 14.12 and 14.14, Equation 14.9 may now be written as

$$\Delta G^{\circ} = \Delta G^{\circ}_{\text{tr}} + \Delta G^{\circ}_{\text{rot}} - RT \ln \left[ (Z^{\text{int}}_{D^* \text{dr}^* \text{aq}} Q_w V) / (Z^{\text{int}}_{D \text{aq}} Z^{\text{int}}_{\text{dr aq}}) \right] + P \Delta V^{\circ}_{\text{aq}} \quad (14.15)$$

Free energy contributions from internal motions that are coupled to solvent are best handled via molecular simulations. Separating the two will amount to an approximation.

$$Z^{\text{int}}_{D \text{aq}} = Z^{\text{intra}}_D Z^{\text{solvn}}_D$$

$$Z^{\text{int}}_{D \text{aq}} \approx \int \dots \int \exp \left[ - \left\{ E(X^N_D) + E(X_D^{\text{Nfixed}}, X^M_W) \right\} / k_B T \right] dX^N_D dX^M_W \\ \approx \left[ \int \dots \int \exp \left\{ -E(X^N_D) / k_B T \right\} dX^N_D \right] \\ \times \left[ \int \dots \int \exp \left\{ -E(X_D^{\text{Nfixed}}, X^M_W) / k_B T \right\} dX^M_W \right] \quad (14.16)$$

Equations similar to Equation 14.16 can be written for D and  $D^* \text{dr}^*$  and converted to excess free energies. Such a separation allows

$$\Delta G^{\circ} = \Delta G^{\circ}_{\text{tr}} + \Delta G^{\circ}_{\text{rot}} + \Delta G^{\circ}_{\text{int}} + \Delta G^{\circ}_{\text{solvn}} \quad (14.17)$$

Equation 14.17 forms the theoretical basis for the additivity assumed in free energy computations as employed in master equation methods [115,116]. The  $P \Delta V^{\circ}_{\text{aq}}$  term in Equation 14.9 is often neglected in liquid-state work. If Equations 14.16 and 14.17 are employed for each structure generated according to Boltzmann distribution either via molecular dynamics or Metropolis Monte Carlo and averages are computed with a suitably calibrated dielectric continuum solvent model for solvation energy for each structure, the results are expected to correspond to Equation 14.9, which is exact (subject to the separability of translations/rotations from internal motions). In inferring free energies of drug–DNA binding from molecular simulations, care must be taken to ensure that the counterions/coions (whether they are “bound” or “released”) are treated as a part of the solute, viz., the DNA or the drug or the complex and their stoichiometries maintained during binding (as in Figure 14.3). Added salt effects are yet to be incorporated in this theory.

$$\Delta G^{\circ}_{\text{int}} = \Delta H^{\circ}_{\text{int}} - T \Delta S^{\circ}_{\text{int}} \quad (14.18)$$

$$\Delta H^{\circ}_{\text{int}} = \Delta H^{\circ}_{\text{intermolecular}} + \Delta H^{\circ}_{\text{intramolecular}}$$

$$\Delta H^{\circ}_{\text{intermolecular}} = \Delta H^{\circ}_{\text{el}} + \Delta H^{\circ}_{\text{vdw}} = \langle \Delta E^{\circ}_{\text{intermolecular}} \rangle = \langle \Delta E^{\circ}_{\text{el}} + \Delta E^{\circ}_{\text{vdw}} \rangle \quad (14.19)$$

$$\Delta H^{\circ}_{\text{intramolecular}} = \langle \Delta E^{\circ}_{\text{intramolecular}} \rangle \quad (14.20)$$

In the above equations,  $\Delta E^{\circ}_{\text{el}}$  and  $\Delta E^{\circ}_{\text{vdw}}$  represent the electrostatic and van der Waals components of the intermolecular interaction energy between the DNA and the drug.

$\Delta E^{\circ}_{\text{intramolecular}}$  represents changes in both bonded and nonbonded contributions to the intramolecular energy of the DNA and the drug upon binding. All these quantities can be computed from a molecular mechanics force field either for a fixed structure (from minimization studies) or for an ensemble of structures from MD simulations.

$$\Delta S^{\circ}_{\text{int}} = \Delta S^{\circ}_{\text{vib,config}} \quad (14.21)$$

Entropy changes can be calculated by a normal mode analysis of an energy-minimized structure ( $\Delta S^{\circ}_{\text{vib}}$ ) or by a quasi-harmonic approximation introduced by Karplus and Kushick [122] and subsequently extended and adapted to MD simulation by Schlitter [123] and van Gunsteren [124]. Equation 14.17 is utilized to generate the energetics of drug–DNA binding from MD simulations.

## REFERENCES

1. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
2. Overington, John P., Al-Lazikani, B. and Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
3. Shaikh, S. A., Jain, T., Sandhu, G., Latha, N. and Jayaram, B. From drug target to leads-sketching, a physicochemical pathway for lead molecule design in silico. *Curr. Pharm. Des.* **13**, 3454–3470 (2007).
4. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. Drugbank: A comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* **34**, 668–672 (2006).
5. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. and Schneider, B. The nucleic acid database: A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759. (1992).
6. Carlton, J. M., Adams, J. H., Silva, J. C., Bidwell, S. L., Lorenzi, H., Caler, E., Crabtree, J. et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, **455**, 757–763 (2008).
7. Young, J. A., Johnson, J. R., Benner, C., Frank Yan, S., Chen, K., Roch, K. G. L., Zhou, Y. and Winzeler, E. A. *In silico* discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics*, **9**, 1–21 (2008).
8. Ptashne, M. A. *Genetic Switch Phage Lambda and Higher Organisms*. Cell Press and Blackwell Scientific, Cambridge, MA. (1992).
9. [http://en.wikipedia.org/wiki/Zinc\\_finger\\_nuclease](http://en.wikipedia.org/wiki/Zinc_finger_nuclease).
10. [http://en.wikipedia.org/wiki/Zinc\\_finger](http://en.wikipedia.org/wiki/Zinc_finger).
11. Liu, Q., Segal, D. J., Ghiara, J. B. and Barbas, C. F. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. USA*, **94**, 5525–5530 (1997).

Q4

12. Kim, J. S. and Pabo, C. O. Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA*. **95**, 2812–2817 (1998).
13. Moore, M., Choo, Y. and Klug, A. Design of polyzinc finger peptides with structured linkers. *Proc. Natl. Acad. Sci. USA*. **98**, 1432–1436 (2001).
14. Moore, M., Klug, A. and Choo, Y. Improved DNA-binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl. Acad. Sci. USA*. **98**, 1437–1441 (2001).
15. Isalan, M., Klug, A. and Choo, Y. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.* **19**, 656–660 (2001).
16. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=CAGE1&search=cage1>.
17. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=BRCA1&search=brca1>.
18. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ABCB1&search=ABCB1>.
19. <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ABCB5&search=ABCB5>.
20. <http://www.genecards.org/index.php?path=/Search/keyword/ABI1>.
21. Lipinski, C. A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today: Technol.* **1**, 337–341 (2004).
22. Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **23**, 3–25 (1997).
23. Hong, M., Fitzgerald, M. X., Harper, S., Luo, C., Speicher, D. W. and Marmorstein, R. Structural basis for dimerization in DNA recognition by Gal4. *Structure*. **16**, 1019–1026 (2008).
24. Bachur N. R., Johnson, R., Yu, F., Hickey, R., Applegren, N., and Malkas, L. Antihelicase action of DNA-binding anticancer agents: Relationship to guanosine-cytidine intercalator binding. *Mol. Pharmacol.* **44**, 1064–1069 (1993).
25. Sun, J. S., Francois, J. C., Garestier, T. M., Behmoaras, T. S., Roig, V., Thuong, N. T. and Helene, C. Sequence-specific intercalating agents: Intercalation at specific sequences on duplex DNA via major groove recognition by oligonucleotide-intercalator conjugates. *Proc. Natl. Acad. Sci. USA*. **86**, 9198–9202 (1989).
26. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **9**, 2215–2235 (2001).
27. Dervan, P. B. and Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **13**, 284–299 (2003).
28. Pilch, D. S., Poklar, N., Baird, E. E., Dervan, P. B. and Breslauer, K. J. The thermodynamics of polyamide-DNA recognition: Hairpin polyamide binding in the minor groove of duplex DNA. *Biochemistry*. **38**, 2143–2151 (1999).
29. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B. and Rees, D. C. A Structural basis for recognition of AzT and TzA base pairs in the minor groove of B-DNA. *Science*. **282**, 111–115 (1998).
30. Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., Dixit, S. et al. A systematic molecular dynamics study of nearest neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* **38**, 299–313 (2010).
31. Beveridge, D. L., Barreiro, G., Byun, K. S., Case, D. A., Cheatham III, T. E., Dixit, S. B., Giudice, E. et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. research design and results on d(CpG) Steps. *Biophys. J.* **87**, 3799–3813 (2004).
32. Shaikh, S. A., Ahmed, S. R. and Jayaram, B. A molecular thermodynamic view of DNA–drug interactions: a case study of 25 minor-groove binders. *Arch. Biochem. Biophys.* **429**, 81–99 (2004).
33. Shaikh, S. A. and Jayaram, B. A swift all-atom energy-based computational protocol to predict DNA–ligand binding affinity. *J. Med. Chem.* **50**, 2240–2244 (2007).

34. Jain, T. and Jayaram, B. An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS Lett.* **579**, 6659–6666 (2005).
35. Arora, N. and Jayaram, B. Energetics of base pairs in B-DNA in solution: An appraisal of potential functions and dielectric treatments. *J. Phys. Chem. B.* **102**, 6139–6144 (1998).
36. Jayaram, B., Swaminathan, S., Beveridge, D. L., Sharp, K. and Honig, B. Monte Carlo simulation studies on the structure of the counterion atmosphere of B-DNA. Variations on the primitive dielectric model. *Macromolecules* **23**, 3156–3165 (1990).
37. M. A. Young, Jayaram, B. and Beveridge, D. L. Local dielectric environment of B-DNA in solution: results from a 14 ns molecular dynamics trajectory, *J. Phys. Chem. B.* **102**, 7666–7669 (1998).
38. Jayaram, B., DiCapua, F. M. and Beveridge, D. L. A theoretical study of polyelectrolyte effects in protein-DNA interactions: Monte Carlo free energy simulations on the ion atmosphere contribution to the thermodynamics of lambda<sub>24</sub> repressor-operator complex formation, *J. Am. Chem. Soc.* **113**, 5211–5215 (1991).
39. Jayaram, B., Sharp, K. A. and Honig, B. The electrostatic potential of B-DNA. *Biopolymers* **28**, 975–993 (1989).
40. Boschtisch, A. H., Fenley, M. O. and Zhou, H.-X. Fast boundary element method for the linear Poisson–Boltzmann equation, *Phys. Chem.* **106**, 2741–2754 (2002).
41. Srinivasan, A. R., Sauers, R. R., Fenley, M. O., Boschtisch, A. H., Matsumoto, A., Colasanti, A. V. and Olson, W. K. Properties of the nucleic-acid bases in free and Watson–Crick hydrogen-bonded states: computational insights into the sequence-dependent features of double-helical DNA. *Biophys. Rev.* **1**, 13–20 (2009).
42. Boschtisch, A. H. and Fenley, M. O. Hybrid boundary element and finite difference method for solving the nonlinear Poisson–Boltzmann equation. *J. Comp. Chem.* **25**, 935–955 (2004).
43. Bredenbergh, J., Boschtisch, A. H. and Fenley, M. O. The role of anionic protein residues on the salt dependence of the binding of aminoacyl-tRNA synthetases to tRNA: A Poisson–Boltzmann analysis. *Commun. Comput. Phys.* **3**, 1051–1070 (2007).
44. Boschtisch, A. H. and Fenley, M. O. A new outer boundary formulation and energy corrections for the nonlinear Poisson–Boltzmann equation. *J. Comp. Chem.* **28**, 909–921 (2007).
45. Fenley, M. O., Harris, R., Jayaram, B. and Boschitsch, A. H. Revisiting the association of cationic groove-binding drugs to DNA using a Poisson–Boltzmann Poisson–Boltzmann approach. *Biophys. J.* **99**, 879–886 (2010).
46. Xu, D., Landon, T., Greenbaum, N. L. and Fenley, M. O. The electrostatic characteristics of GU wobble base pairs. *Nucleic Acids Res.* **35**, 3836–3847 (2007).
47. Xu, D., Greenbaum, N. L. and Fenley, M. O. Recognition of the spliceosomal branch site RNA helix on the basis of surface and electrostatic features. *Nucleic Acids Res.* **33**, 1154–1161 (2005).
48. Bredenbergh, J. H., Russo, C. and Fenley, M. O. Salt-mediated electrostatics in the association of TATA binding proteins to DNA: A combined molecular mechanics/Poisson–Boltzmann study. *Biophys. J.* **94**, 4634–4645 (2008).
49. Jen-Jacobson, L. Protein–DNA recognition complexes: Conservation of structure and binding energy in the transition state, *Biopolymers* **44**, 153 (1997).
50. Pineda De Castro, L. F. and Zacharias, M. DAPI binding to the DNA minor groove: A continuum solvent analysis. *J. Mol. Recogn.* **15**, 209–220 (2002).
51. Vibrational stark effect spectroscopy at the interface of Ras and Rap1A bound to the Ras binding domain of RalGDS reveals an electrostatic mechanism for protein-protein interactions, *J. Phys. Chem. B* (2010), appeared online.
52. Bertonati, C., Honig, B. and Alexov, E. Poisson–Boltzmann Poisson–Boltzmann calculations of nonspecific salt effects on protein-protein binding free energies. *Biophys. J.* **92**, 1891–1899 (2007).

53. Anderson, C. F. and Record, M. T. J. Salt dependence of oligoion–polyion binding: A thermodynamic description based on preferential interaction coefficients. *J. Phys. Chem.* **97**, 7116–7126 (1993).
54. Manning, G. S. The molecular theory of polyelectrolyte solutions with application to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11**, 179–246. (1978).
55. Record, M. T. J., Anderson, C. F. and Lohman, T. M. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.* **11**, 103–178 (1978).
56. Record, M. T. J., Zhang, W. and Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, Hogmeister effects and osmotic effects of salts. *Adv. Prot. Chem.* **51**, 281–353 (1998).
57. Chaires, J. B. Calorimetry and thermodynamics in drug design. *Annu. Rev. Biophys.* **37**, 135–151 (2008).
58. Rohs, R., and Sklenar, H. Methylene blue binding to DNA with alternating AT base sequence: Minor groove binding is favored over intercalation. *J. Biomol. Struct. Dyn.* **21**, 399–711 (2004).
59. Sharp, K. A. Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson–Boltzmann model. *Biopolymers.* **36**, 227–243 (1995).
60. Misra, V. K. and Honig, B. On the magnitude of the electrostatic contribution to ligand–DNA interactions. *Proc. Natl. Acad. Sci.* **92**, 4691–4695 (1995).
61. Misra, V. K., Sharp, K. A., Friedman, R. A. and Honig, B. Salt effects on ligand–DNA binding: Minor groove binding antibiotics. *J. Mol. Biol.* **238**, 245–263 (1994).
62. Sharp, K. A., Friedman, R. A., Misra, V., Hecht, J. and Honig, B. Salt effects on polyelectrolyte–ligand binding: Comparison of Poisson–Boltzmann, and limiting law/counterion binding models. *Biopolymers.* **36**, 245–262 (1995).
63. Baginski, M., Fogolari, F. and Briggs, J. M. Electrostatic and non-electrostatic contributions to the binding free energies of anthracycline antibiotics to DNA. *J. Mol. Biol.* **274**, 253–267 (1997).
64. Chen, S.-W. and Honig, B. Monovalent & divalent salt effects on electrostatic free energies defined by the nonlinear Poisson–Boltzmann equation: Application to DNA binding reactions. *J. Phys. Chem.* **101**, 9113–9118 (1997).
65. Rohs, R., Sklenar, H., Lavery, R. and Roder, B. Methylene blue binding to DNA with alternating GC base sequence: A modeling study. *J. Am. Chem. Soc.* **122**, 2860–2866 (2000).
66. Misra, V. K. and Honig, B. On the magnitude of the electrostatic contribution to ligand–DNA interactions. *Proc. Nat. Acad. Sci. USA.* **92**, 4691–4695 (1995).
67. Strekowski, L. and Wilson, B. Noncovalent interactions with DNA: An overview. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **623**, 3–13. (2007).
68. Wilson, W. D., Tanious, F. A., Mathis, A., Tevis, D., Hall, J. E. and Boykin, D. W. Antiparasitic compounds that target DNA. *Biochimie.* **90**, 999–1014 (2008).
69. Lane, A. N. and Jenkins, T. C. Thermodynamics of nucleic acids and their interactions with ligands. *Q. Rev. Biophys.* **33**, 255–306 (2000).
70. Mazur, S., Tanious, F. A., Ding, D., Kumar, A., Boykin, D. W., Simpson, I. J., Neidle, S. and Wilson, W. D. A thermodynamic and structural analysis of DNA minor-groove complex formation. *J. Mol. Biol.* **300**, 321–337 (2000).
71. Nguyen, B., Lee, M. P. H., Hamelberg, D., Joubert, A., Bailly, C., Brun, R., Neidle, S. and Wilson, W. D. Strong binding in the DNA minor groove by an aromatic diamidine with a shape that does not match the curvature of the groove. *J. Am. Chem. Soc.* **124**, 13680–13681 (2002).

72. Miao, Y., Lee, M. P. H., Parkinson, G. N., Batista-Parra, A., Ismail, M. A., Neidle, S., Boykin, D. W. and Wilson, W. D. Out-of-shape DNA minor groove binders: Induced fit interactions of heterocyclic dications with the DNA minor groove. *Biochemistry*. **44**, 14701–14708 (2005).
73. Chaires, J. B. Energetics of drug-DNA interactions. *Biopolymers*. **44**, 201–215 (1998).
74. Lin, P.-H., Yen, S.-L., Lin, M.-S., Chang, Y., Louis, S. R., Higuchi, A. and Chen, W.-Y. Microcalorimetric studies of the thermodynamics and binding mechanism between L-tyrosinamide and aptamer. *J. Phys. Chem.* **112**, 6665–6673 (2008).
75. Hossain, M., and Kumar, G. S. DNA binding of benzophenanthridine compounds sanguinarine versus ethidium: Comparative binding and thermodynamic profile of intercalation. *J. Chem. Thermodynamics*. **41**, 764–774 (2009).
76. Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S. and Honig, B. The role of DNA shape in protein-DNA recognition. *Nature*. **461**, 1248–1253 (2009).
77. Eargle, T. J., Black, A. A., Sethi, A., Trabuco, L. G. and Luthey-Schulten, Z. Dynamics of recognition between tRNA and elongation factor. *J. Mol. Biol.* **377**, 1382–1405 (2008).
78. Paleskava, S. A., Knevega, A. L. and Rodnina, M. V. Thermodynamic and kinetic framework of selenocysteyl-tRNA<sup>Sec</sup> recognition by elongation factor. *J. Biol. Chem.* **285**, 3014–3020 (2010).
79. Spiegel, K. and Magistrato, A. Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations. *Org. Biomol. Chem.* **4**, 2507–2517 (2006).
80. Dolenc, J., Oostenbrink, C., Koller, J., and Gunsteren, W. F. van. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site. *Nucleic Acids Res.* **33**, 725–733 (2005).
81. Orzechowski, M. and Cieplak, P. Application of steered molecular dynamics (SMD) to study DNA–drug complexes and probing helical propensity of amino acids. *J. Phys. Condens. Matter*. **17**, 1627 (2005).
82. Magistrato, A., Ruggerone, P., Spiegel, K., Carloni, P. and Reedijk, J. Binding of novel azole-bridged dinuclear platinum(II) anticancer drugs to DNA: Insights from hybrid QM/MM molecular dynamics simulations. *J. Phys. Chem. B*. **110**, 3604–3613 (2006).
83. Ma, G. Z., Zheng, K. W., Jiang, Y. J., Zhao, W. N., Zhuang, S. L. and Yu, Q. S. Minor groove binding between norfloxacin and DNA duplexes in solution: A molecular dynamics study. *Chinese Chem. Lett.* **16**, 1367–1370 (2005).
84. Tajmir-Riahi, H. A. AZT binding to DNA and RNA: Molecular modeling and biological significance. *J. Iranian Chem. Soc.* **2**, 78–84 (2005).
85. Spiegel, K., Rothlisberger, U., and Carloni, P. Duocarmycins binding to DNA investigated by molecular simulation. *J. Phys. Chem. B*. **110**, 3647–3660 (2006).
86. Vargiu, A. V., Ruggerone, P., Magistrato, A. and Carloni, P. Anthramycin–DNA binding explored by molecular simulations. *J. Phys. Chem. B*. **110**, 24687–24695 (2006).
87. Fang, Y. Morris, V. R., Lingani, G. M., Long, E. C. and Southerland, W. M. Genome-targeted drug design: Understanding the netropsin–DNA interaction. *Open Conf. Proc. J.* **1**, 157–163 (2010).
88. Marco, E. and Gago, F. DNA structural similarity in the 2:1 complexes of the antitumor drugs trabectedin (yondelis) and chromomycin A3 with an oligonucleotide sequence containing two adjacent TGG binding sites on opposing strands. *Mol. Pharmacol.* **68**, 1559–1567 (2005).
89. Still, W. C., Tempczyk, A., Hawley, R. C. and Hendrickson, T. J. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
90. Jayaram, B., Liu, Y. and Beveridge, D. L. A modification of the generalized Born theory for improved estimates of solvation energies and pKa shifts. *J. Chem. Phys.* **109**, 1465–1471 (1998).

91. Hawkins, G. D., Cramer, C. J. and Truhlar, D.G. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824–19839 (1996).
92. Jayaram, B., Sprous, D., and Beveridge, D. L. Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field. *J. Phys. Chem. B.* **102**, 9571–9576 (1998).
93. Jayaram, B., McConnell, K., Dixit, S. B., Das, A. and Beveridge, D. L. Free-Energy component analysis of 40 protein-DNA complexes: A consensus view on the thermodynamics of binding at the molecular level. *J. Comput. Chem.* **23**, 1–14 (2002).
94. Jayaram, B. and Jain, T. The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 343–61 (2004).
95. Lavery, R. and Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91 (1988).
96. Lavery, R. and Sklenar, H. Defining the structure of irregular nucleic acids: Conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655–667 (1989).
97. Bhattacharya, D. and Bansal, M. A general procedure for generation of curved DNA molecules. *J. Biomol. Struct. Dyn.* **6**, 93–104 (1988).
98. Bhattacharya, D. and Bansal M. A self-consistent formulation for analysis and generation of non-uniform DNA structures. *J. Biomol. Struct. Dyn.* **6**, 635–653 (1989).
99. Bhattacharyya, D. and Bansal, M. Local variability and base sequence effects in DNA crystal structures. *J. Biomol. Struct. Dyn.* **8**, 539–572 (1990).
100. Bhattacharya, D. and Bansal, M. Groove width and depths of B-DNA structures depend on local variation in slide. *J. Biomol. Struct. Dyn.* **10**, 213–226 (1992).
101. Bansal, M., Bhattacharya, D. and Ravi B. NUPARM and NUCGEN: Software for analysis and generation of sequence dependent nucleic acid structures. *Comput. Appl. Biosci.* **11**, 281–287 (1995).
102. Mukherjee, S., Bansal, M. and Bhattacharyya, D. Conformational specificity of non-canonical base pairs 4 and higher order structures in nucleic acids: crystal 5 structure database analysis. *J. Comp. Aided Mol. Des.* **20**, 629–645 (2006).
103. Lu, X. J. and Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res.* **31**, 5108–5121 (2003).
104. Lu, X. J. and Olson, W. K. “3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures,” *Nat. Protoc.* **3**, 1213–27 (2008).
105. Case, D. A., Cheatham, III, T.E., Darden, T., Gohlke, H., Luo, R., Merz, Jr., K.M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. The Amber biomolecular simulation programs. *J. Computat. Chem.* **26**, 1668–1688 (2005).
106. Ponder, J. W. and Case, D. A. Force fields for protein simulations. *Adv. Prot. Chem.* **66**, 27–85 (2003).
107. Arnott, S. Campbell-Smith, P.J. and Chandrasekaran, R. In Fasman, Ed. *Handbook of Biochemistry and Molecular Biology*, 3rd ed. Nucleic Acids, Vol. II, G.P. Cleveland: CRC Press (1976). pp. 411–422.
108. *Surflex, version 2.11*. Tripos, Inc.: St. Louis, MO 2007.
109. *Autodock, version 4*. The Scripps Research Institute: La Jolla, CA, 2007.
110. Gupta, A., Gandhimathi, A., Sharma, P., and Jayaram, B. (2007) ParDOCK: An all atom energy based Monte Carlo docking protocol for protein–ligand complexes. *Protein Pept. Lett.* **14**, 632–646 (2007).
111. Jones, G., Willett P. and Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **245**, 43–53 (1995).
112. Jones, G., Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).

113. Nissink, J. W. M., Murray, C., Hartshorn, M., Verdonk, M. L., Cole, J. C. and Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins*. **49**, 457–471 (2002).
114. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. and Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins*. **52**, 609–623 (2003).
115. Cole, J. C., Nissink, J. W. M. and Taylor, R. In B. Shoichet, J. Alvarez Eds. Protein-ligand docking and virtual screening with GOLD. *Virtual Screening in Drug Discovery*, Taylor & Francis, Boca Raton, Florida, USA (2005).
116. Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W. M., Taylor, R. D., and Taylor, R. Modeling water molecules in protein-ligand docking using GOLD *J. Med. Chem.* **48**, 6504–6515 (2005).
117. Hartshorn, M. J., Verdonk, M. L., Chessari, G., Brewerton, S. C., Mooij, W. T. M., Mortenson, P. N. and Murray, C. W. High-quality test set for the validation of protein-ligand docking performance *J. Med. Chem.* **50**, 726–741 (2007).
118. Dominguez, C., Boelens, R. and Bonvin, A. M.J.J. HADDOCK: A protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
119. Ausiello, G., Cesareni, G. and Helmer Citterich, M. ESCHER: A new docking procedure applied to the reconstruction of protein tertiary structure, *Proteins*. **28**, 556–567 (1997).
120. Jayaram, B., McConnell, K. J., Dixit, S. B. and Beveridge, D. L. Free energy analysis of protein-DNA binding: The *ecoRI* endonuclease—DNA complex. *J. Comput. Phy.* **151**, 333–357 (1999).
121. Kalra, P., Reddy, T.V. and Jayaram, B. Free energy component analysis for drug design: A case study of HIV-1 protease-inhibitor binding. *J. Med. Chem.* **44**, 4325–4338 (2001).
122. Karplus, M. and Kaushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules*. **14**, 325–332 (1981).
123. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the Covariance matrix. *Chem. Phys. Lett.* **215**, 617–621 (1993).
124. Schafer, H., Mark, A. E. and van Gunsteren. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* **113**, 7809–7817 (2000).

**TO: CORRESPONDING AUTHOR****AUTHOR QUERIES – TO BE ANSWERED BY THE AUTHOR**

The following queries have arisen during the typesetting of your manuscript. Please answer these queries by marking the required corrections at the appropriate point in the text.

<b>Query No.</b>	<b>Query</b>	<b>Response</b>
Q1	Please provide the expansion for RCSB.	
Q2	Please check if the citation of Figure 14.3 is OK here.	
Q3	Please check if $K_{eq,aq}$ is OK.	
Q4	The following references are not cited in the text: [2,3,4,5,16,17,18,19,20,95–119]. Please cite them.	
Q5	Please provide author names, volume no and page range in Ref. 51.	
Q6	Are “PreDDICTA” and “preDDicta” two different methods?	