J. Mol. Biol. (1998) **276**, 57–70

JMB

A Theoretical Analysis of Specificity of Nucleic Acid Interactions with Oligonucleotides and Peptide Nucleic Acids (PNAs)

Aleksey Lomakin[1] and Maxim D. Frank-Kamenetskii[2]*

[1]*Physics Department Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[2]*Center for Advanced Biotechnology and Department of Biomedical Engineering Boston University 36 Cummington St., Boston MA 02215, USA*

**Corresponding author*

We treat theoretically the problem of the specificity of interaction between nucleic acid and an oligonucleotide, its analog or its mimic (such as peptide nucleic acid, or PNA). We consider simplest models with only essential details using numerical solutions of kinetic equations and the kinetic Monte Carlo method. In our first model, describing the formation of complementary duplex, we demonstrate anti-correlation between specificity and affinity for nucleic acid/oligonucleotide interaction.

We analyze in detail one notable exception. Homopyrimidine PNAs exhibit very high affinity to DNA forming extraordinarily stable DNA/$(PNA)_2$ triplexes with the complementary DNA strand. At the same time, such PNAs show remarkable sequence specificity of binding to duplex DNA. We formulate a theoretical model for the two-step process of PNA interaction with DNA. The calculations demonstrate that two-stage binding may secure both high affinity and very high specificity of PNA interaction with DNA. Our computer simulations define the range of parameter values in which high specificity is achieved. These findings are of great importance for numerous applications of PNA and for design of future drugs which specifically interact with DNA.

Keywords: nucleic acids; specificity of hybridization; affinity of interaction; PNA/DNA complexes; triplexes

Introduction

Problem of specificity of biomolecular interaction

The present theoretical study elucidates the relationship between specificity and affinity of interaction between nucleic acids, their analogs and mimics. As usual, we refer to cases when the equilibrium binding constant has a high value as high affinity situations, and we refer to cases where there is an efficient discrimination between specific complexes and non-specific ones as high specificity situations.

The question of specificity of biomolecular assembly is related to virtually all problems of biology and medicine. Specific interactions between nucleic acids, between nucleic acids and proteins, between proteins, between proteins and their ligands play a pivotal role in a large variety of biological processes. In medicine, the theme is central for many areas, from immunology to drug design. In spite of its paramount significance, the basic thermodynamic and kinetic principles of the specific recognition are not well understood. Specificity of biological processes has long been taken for granted and has been considered as an inherent property of living systems. Only recently with the advent of protein engineering, combinatorial chemistry, the aptamer technology and in connection with rational drug design, a comprehensive analysis of the problem has been initiated (von Hippel & Berg, 1986; von Hippel, 1994; Spolar & Record, 1994; Eaton *et al*., 1995). It has become obvious that progress in the direction of rational drug design critically depends on our understanding of the basic principles of specificity of biomolecular recognition.

The problem may seem a very simple one, at least in principle. Indeed, following von Hippel & Berg (1986) (we will refer to this seminal paper as vHB) the specificity, R, may be defined as a ratio

Abbreviations used: TFO, triplex forming oligonucleotides; PNA, peptide nucleic acid; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA.

0022–2836/98/060057–14 $25.00/0/mb971497

of concentration, $[TL]$, of complexes of the ligand, L, with its specific target, T, and the sum of concentrations of complexes, $[T_iL]$, of the ligand with all non-specific sites T_i:

$$R = [TL]/\Sigma[T_iL] \quad (1)$$

Simple arguments based on the mass action law then lead to the final equation for R (von Hippel & Berg, 1986):

$$R = K[T]/\Sigma K_i[T_i] \quad (2)$$

where K and K_i are the equilibrium constants of binding of the ligand to the specific and ith non-specific sites, respectively.

Equation (2) leads to the intuitively obvious conclusion that high specificity is achieved when K is much larger than all K_i taken together (under the natural assumption that $[T]$ and each of $[T_i]$ are of the same order of magnitude). For the sake of simplicity and without losing generality, we discuss throughout the present work only two types of complexes: specific, also referred to as correct or matched and non-specific, also referred to as frustrated or mismatched. A mismatched complex is actually a representative of a large number of possible non-specific complexes. Within the framework of the von Hippel & Berg (1986) treatment, the matched and mismatched complexes are characterized by two equilibrium binding constants, K and $\hat{K}$, respectively. High specificity requires a strong inequality to be valid:

$$K \gg \hat{K} \quad (3)$$

Thus, within the framework of the von Hippel & Berg (1986) analysis, high selectivity of binding may only be the result of a huge free energy gap between the correct and mismatched or frustrated complexes. On the basis of the vHB-type analysis, Eaton *et al*. (1995) formulated an important principle of drug selection, according to which high affinity binding (i.e. binding with a high value of the equilibrium binding constant) entails highly specific binding.

The above considerations explain, at least qualitatively, the remarkable selectivity of biomolecular interactions in a variety of real situations, both in nature and in the course of selection of new drugs (such as aptamers: von Hippel & Berg, 1986; von Hippel, 1994; Spolar & Record, 1994; Eaton *et al*., 1995). However, this reasoning uses two essential assumptions: (i) the concentrations of ligands bound to various targets achieve equilibrium, and (ii) the strong inequality in equation (3) is valid for any non-specific target. These rather restrictive assumptions are not universally valid and therefore vHB-type treatment is sometimes inapplicable and its conclusions fail.

The specificity of interaction between nucleic acids plays a crucial role in fundamental biological processes such as replication, transcription, translation and genetic recombination. The ability of single-stranded nucleic acids to form specific complexes is crucial for most important biotechnological applications of nucleic acids, such as various hybridization techniques (Southern and Northern blotting, *in situ* hybridization, polymerase chain reaction (PCR), the antisense strategy and others). Recognition of nucleic acids stands out among other biomolecular recognition processes in that it is achieved by a remarkably simple general mechanism, the Watson–Crick and/or Hoogsteen pairing. As a result, a comprehensive theoretical treatment of nucleic acid interactions is possible within the framework of very simple models. In these models the real interactions could be stripped of all unnecessary details and only the features important for understanding of the specificity of interaction are left. The specificity of PNA/dsDNA interaction presents a special and most interesting case, which we study in detail here.

An artificial oligonucleotide mimic, peptide nucleic acid (PNA), consists of a protein-like backbone and ordinary DNA bases (Nielsen *et al*., 1991). It has been found to form stable and sequence-specific complexes with single-stranded and duplex DNA. PNA shows stronger affinity to DNA than ordinary oligonucleotides. As a result, PNA finds more and more applications (Hyrup & Nielsen, 1996; Dueholm & Nielsen, 1997). PNA is especially promising in sequence-specific targeting of dsDNA. Homopyrimidine PNAs form very unusual complexes with complementary sites on dsDNA. In these complexes, two PNA oligomers form a triplex with the homopurine strand of DNA leaving the other DNA strand displaced (see Figure 1). Binding of PNA to dsDNA exhibits a remarkable combination of high specificity and affinity (Demidov *et al*., 1995, 1996; Veselkov *et al*., 1996a,b; Kuhn *et al*., 1998; Demidov *et al*., unpublished results). In contrast to canonical nucleic acid interactions, the basic facts concerning PNA/DNA interaction are not widely known. What follows is a concise overview of the basic experimental facts concerning the interaction of PNA with DNA.

Complexes of PNA with DNA

PNA oligomers with mixed purine-pyrimidine composition form duplexes with complementary PNA and ssDNA chains (Egholm *et al*., 1993; Wittung *et al*., 1994). The melting temperature of PNA/PNA duplexes is independent of the ionic strength (Tomac *et al*., 1996), whereas the melting temperature of PNA/ssDNA duplexes decreases slightly with increasing ionic strength. By contrast, the melting temperature of DNA/DNA duplexes is well known to increase with increasing ionic strength. Tomac *et al*. (1996) pointed out the neutrality of the PNA backbone to quantitatively explain the peculiar behavior of the PNA/ssDNA duplexes using the polyelectrolyte theory of Frank-Kamenetskii *et al*., (1987) and Bond *et al*., (1994). Although more stable, PNA/ssDNA duplexes are not particularly different from DNA/DNA duplexes in other respects.

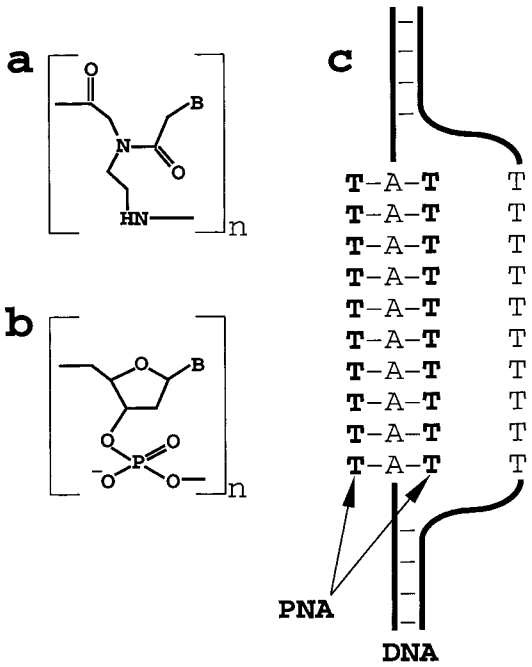

Figure 1. Chemical structure of (a) PNA, and (b) DNA, and schematics of complex between duplex DNA and homopyrimidine PNA (decathymine-PNA as an example) c, Such a complex is often called a P-loop.

The difference between PNA and DNA oligomers manifests itself most strikingly in their interaction with dsDNA. Only homopyrimidine PNAs are known to form stable complexes with dsDNA. The structure of these triplexes is completely different from $(DNA)_3$ triplexes formed between homopyrimidine DNA oligomers and complementary sites on dsDNA (Frank-Kamenetskii & Mirkin, 1995; Soyfer & Potaman, 1996). As Figure 1 shows, two homopyrimidine PNA oligomers invade dsDNA forming a so-called P-loop (Cherny, 1993; Nielsen *et al*., 1994; Demidov *et al*., 1995).

A major element of the P-loop is the $(PNA)_2$/DNA triplex. The P-loop forms in spite of the unfavorable helix boundaries only because this triplex is remarkably stable. Two factors contribute to its extraordinary stability. First, the neutrality of PNA eliminates the strong electrostatic repulsion between the strands characteristic of the $(DNA)_3$ triplex. Second, according to X-ray crystallographic data of Betts *et al*. (1995), in addition to Watson–Crick and Hoogsteen pairing, the $(PNA)_2$/DNA triplex is stabilized by hydrogen bonds between amide (peptide) nitrogen atoms of the Hoogsteen-bonding PNA strand and the phosphate oxygen atoms of the DNA backbone. Because two PNA oligomers are involved in the recognition process, PNA clamps or bis-PNAs, which carry two PNA oligomers connected by a flexible linker, are often used to target homopurine sites on nucleic acids (Egholm *et al*., 1995; Griffith *et al*., 1995; Hyrup & Nielsen, 1996; Dueholm & Nielsen, 1997). Additional stabilization of the complex could be achieved by incorporating positive charges into PNA or bis-PNA oligomers (Griffith *et al*., 1995; Hyrup & Nielsen, 1996; Demidov *et al*., 1996; Veselkov *et al*., 1996a,b; Dueholm & Nielsen, 1997; Kuhn *et al*., 1998).

The additional hydrogen bonds, and the absence of electrostatic repulsion between the strands make the $(PNA)_2$/DNA triplexes the most stable nucleic acid-like complexes known to date. It has been suggested that the formation of $(PNA)_2$/DNA triplexes occurs in a two-step process (Demidov *et al*., 1995, 1996). We demonstrate here how in a two-stage binding a very high specificity of the PNA interaction with dsDNA can be secured. These results may have implications for the future design of PNA-based drugs or other DNA mimics.

Models

Model no. 1: one-step binding

Let us first formulate our model for the case of a perfect match between the ligand and its target site. The ligand and the target site are complementary chains consisting of m residues. The chains may be separated or form complexes of different types. We assume that each residue in the target can be either free or linked exclusively with its counterpart in the ligand. Only consecutive residues can be linked. Any state of the complex may then be fully characterized by the domain of linked residues. There is one free, unlinked state, m states with one link, $m-1$ states with two consecutive links and so on up to the state with all residues linked. All m processes corresponding to the transition between free state and the states with one link (Figure 2A) are characterized by the same values of forward and backward kinetic constants, k_b^+ and k_r .

The processes within the bound state are shown in Figure 2B. All these processes have identical elongation kinetic constants k_e. The reduction kinetic constants k_r are also identical, including one for the very last link in Figure 2A. The equilibrium stability constant for the duplex elongation is defined as $s = k_e/k_r$. Constants k_e and s are sufficient to describe all elementary kinetic steps for any m. To introduce mismatched pairing, one only needs to ascribe a different value $\hat{k}_r$ for the reduction kinetic constants of a residue. We will characterize mismatched links by a frustration factor f, so that $\hat{k}_r = fk_r$ or for the stability constant, $\hat{s} = s/f$.

Model no. 1 is a simplified version of the models describing melting and association of complementary strands of nucleic acids. We keep here only essential features, and ignore the possibility of staggered binding, multiple nucleation and other. This is very similar to a model we used earlier to treat the slow relaxation processes in DNA melting (Anshelevich *et al*., 1984). In that work the problem of strand dissociation was treated analytically for

chains of arbitrary length. Here, we consider only short chains (consisting of about ten residues), which makes it possible to apply numerical methods to treat both association and dissociation of strands rigorously and to incorporate new elements into the model, such as triplex formation in the case of complexing DNA with PNA. It should be emphasized that numerical methods we use here are restricted to short chains and cannot replace the analytical treatment of strand dissociation for long chains.

We will use Model no. 1 to better understand the specificity/affinity relationship for two important real cases: (i) targeting of single-stranded nucleic acids with oligonucleotides *via* Watson–Crick pairing, and (ii) targeting of duplex DNA with triplex-forming oligonucleotides (TFO) *via* Hoogsteen pairing (Frank-Kamenetskii & Mirkin, 1995; Soyfer & Potaman, 1996). This model is equally applicable to the case of targeting single-stranded nucleic acids with PNA oligomers of mixed purine-pyrimidine composition. Like oligonucleotides, such PNAs also form complexes with single-stranded nucleic acids *via* Watson–Crick pairing. However, the model needs to be substantially modified to be applicable for the case of targeting both single-stranded and duplex DNA with homopyrimidine PNAs.

Model no. 2: two-step binding

Homopyrimidine PNA oligomers form triplexes with complementary single-stranded nucleic acid targets. These triplexes are so exceptionally stable that PNA oligomers containing as few as five to seven residues target dsDNA. In the resulting P-loop structure the DNA strand having the same sequence as the PNA oligomer is displaced while the purine DNA strand forms a triplex with two PNA oligomers. A kinetic model has been proposed according to which the P-loop is formed at two stages: at the first stage a transient Watson–Crick DNA/PNA duplex is formed while at the second stage the second PNA oligomer is attached forming the triplex (Demidov *et al*., 1995, 1996). Alternatively, the first stage of the process may consist of a transient formation of a triplex between duplex DNA and one PNA oligomer *via* Hoogsteen pairing (Demidov *et al*., 1995; Wittung *et al*., 1996). At the second stage, the second PNA oligomer is attached displacing one DNA strand thereby replacing the $(DNA)_2$/PNA triplex with the $(PNA)_2$/DNA triplex (i.e. forming the P-loop). The net result of both mechanisms is the same and it is still not clear *via* which particular route the process of formation of the P-loop proceeds. Both mechanisms, however, are two-stage processes, which makes the formation of the complex between duplex DNA and homopyrimidine PNA fundamentally different from the DNA/DNA duplex formation and the duplex DNA/oligonucleotide triplex formation. A main goal of our work is to show that such a two-stage binding may exhibit specificity/affinity features quite different from those in a single-stage mode of binding described by Model no. 1.

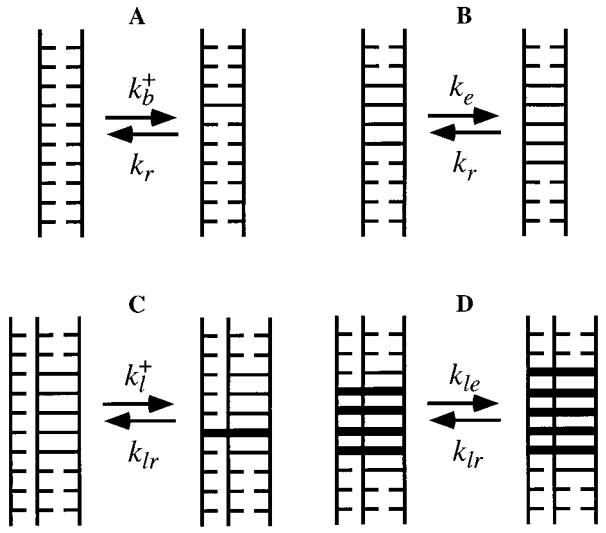

Figure 2. Simplified kinetic scheme of one-step (Model no. 1) and two-step (Model no. 2) binding. Model no. 1 postulates initial binding of any site with the rate k_b^+ (A), and subsequent elongation of the bound domain with the rate k_e (B). Reduction of links, including the last one, occurs with the rate k_r only at endpoints of the bound domain. Model no. 2 includes an additional ''locking'' stage. Any link within the bound domain can be locked with the rate $\bar{k}_l^+$, so that the overall rate of the initiation of locking of the state with n links is $k_l^+ = n \cdot \bar{k}_l^+$. C, Elongation of the locked region occurs only within the bound domain with the rate k_{le} D. ''Unlocking'' occurs only at the endpoints of the locked region with the rate k_{lr}. (See additional explanations in the text.)

The two-stage binding is incorporated into Model no. 2. Within the framework of this model, there are two, rather than one, ligands. The first ligand binds to the target according to Model no. 1 (Figure 2A and B). The second ligand may form complexes only with the duplex part of the complex between the target and the first ligand. Thus, in our new model, each position may assume three states: no link, duplex and triplex. After the appearance of the first duplex bond, a new set of events may happen, which correspond to the nucleation of the triplex (Figure 2C). Since the triplex is extraordinarily stable we will refer to this process as ''locking''. The corresponding locking rate is k_l^+. After locking has occurred the triplex domain can grow, but only where the duplex is already formed (Figure 2D). The elongation of the triplex domain is associated with kinetic constant k_{le} and its reduction with constant k_{lr} is assumed to be the same for all triplex links including the reduction of the last link between the duplex and the second ligand (Figure 2C). The equilibrium stability constant for the triplex elongation is defined as $s_l = k_{le}/k_{lr}$, and for the sake of simplicity

we assume that the mismatch in the duplex does not affect either the locking rate at this position or k_{le} and k_{lr}.

Model no. 2 is equally applicable to the homopyrimidine PNA binding to a single-stranded DNA as well as to a double-stranded DNA. In the case of single-stranded DNA the duplex between target and first ligand in Figure 2 is formed by Watson–Crick pairing between DNA and PNA. In the case of double-stranded DNA though, the interpretation of the ''duplex'' in Figure 2 is not that straightforward. According to Demidov *et al*. (1995) and Wittung *et al*. (1996), the first stage of binding between PNA and duplex DNA could be either opening of the DNA double helix and formation of the Watson–Crick duplex between one PNA oligomer and the complementary DNA strand (the ''Watson–Crick-first'' mechanism) or formation of highly unstable $(DNA)_2$/PNA triplex *via* Hoogsteen pairing (the ''Hoogsteen-first'' mechanism). The kinetics of the overall process is described by Model no. 2 in both cases. The interpretation of the kinetic constants, however, depends on the actual mechanism of reaction.

Calculation Methods

Monte Carlo

We use a standard kinetic Monte Carlo procedure (Hammersley & Handscamb, 1964; Lukashin *et al*., 1976). In this procedure an instance of kinetic evolution of the system is generated in the following way. Consider the system in a state α at the moment t. The time t_α for the system to remain in the state α is given by:

$$t_\alpha = \ln(\gamma)/\Sigma k_{\alpha i}$$

where γ is a random number uniformly distributed over the [0,1] interval, and $k_{\alpha i}$ are the kinetic constants for the transitions from the state α to any other possible state i. The particular state β to which the transition occurs at the moment $t + t_\alpha$ is to be chosen with the probability $k_{\alpha\beta}/\Sigma k_{\alpha i}$. Then the whole procedure is repeated with the new set of kinetic constants for the state β.

Kinetic equations

A system of linear kinetic equations corresponding to our models can be written. The number of kinetic equations is equal to the total number of states in the system. For Model no. 1 it is $m(m+1)/2 + 1$. This number is sufficiently small for the values of m we consider here to allow the direct diagonalization of the matrix of kinetic coefficients. In the case of Model no. 2, the total number of states is very large (about $m^4/24$), which makes the diagonalization method impractical. However, in the experimentally interesting case of virtually irreversible triplex formation, all states with triplexes in Model no. 2 are in partial equilibrium with each other and can be considered as one state, which we dub ''locked''. Thus, the total number of states of interest in Model no. 2 with sufficiently stable triplex is larger by 1 than the number of states in Model no. 1. In this case the diagonalization method is again applicable. It is much faster than the direct Monte Carlo simulations especially for small rates of initial binding and/or locking. These methods yielded identical results, within the statistical error of the Monte Carlo method.

Parameter values

As will become clear later, without loss of generality we may assume that elongation constants for binding and locking are the same, and measure time in units $1/k_e$, i.e. we take $k_e = k_{le} = 1$. The kinetic constants k_b^+ and k_l^+, the stability constants $s = k_e/k_r$ and $s_l = k_{le}/k_{lr}$, the frustration parameter f, the number of binding sites in the target sites in the target and oligonucleotide m and the position of mismatches form the complete set of parameters describing Model no. 2. Model no. 1 is simply a special case of Model no. 2 with $k_l^+ \to 0$. Values of all parameters listed above significantly affect the kinetics of the reaction.

The parameters k_b^+ and k_l^+ are proportional to the ligand concentration. (In the case of bis-PNAs k_l^+ is independent of PNA concentration, but it varies widely with the type of linker between PNA oligomers, see Kuhn *et al*., 1998). We will vary k_l^+ to check how it affects the relationship between specificity and affinity. Both these constants should be small to affect kinetics in a non-trivial way, k_l^+, $k_b^+ \ll 1$. Stability constants s and s_l may vary between 1 and about 10 depending on ambient conditions (temperature, salt concentration, pH). Note that s_l is usually larger than s because of remarkable stability of the $DNA/(PNA)_2$ triplex. The value of frustration parameter f depends on the type of mismatch and is typically of the order of 10^2 to 10^3 (Tibanyenda *et al*., 1984; Egholm *et al*., 1993; Wittung *et al*., 1994; Demidov *et al*., 1995; Kuhn *et al*., 1998). For such a big f, however, the time scales for matched and mismatched binding are too different, and we will focus our attention on the domain $f = 10$ to 100 where the comparison between matched and mismatched kinetics is more vivid. Unless explicitly stated otherwise our calculations are made for $m = 10$ and a single mismatch located at the fifth position.

Results and Discussion

Specificity of recognition *via* Watson–Crick pairing

We start by considering the case of targeting the single-stranded DNA or RNA by the DNA or RNA oligonucleotide. In this case, which falls under Model no. 1, there are only two essential parameters, k_b^+ and s. The binding rate k_b^+ depends mostly on the oligonucleotide concentration,

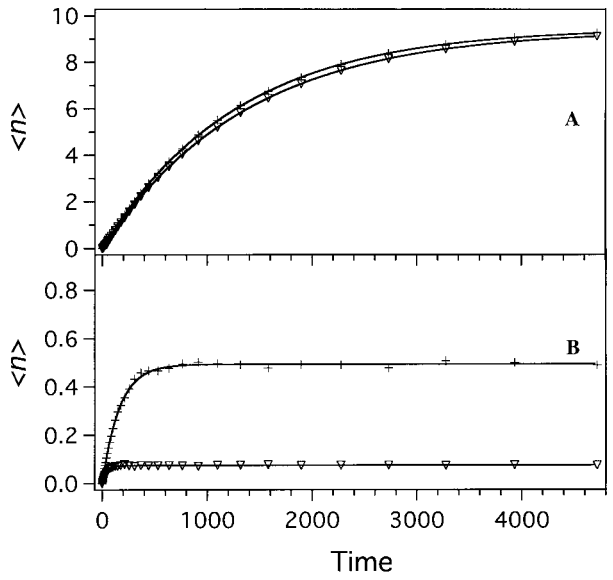

Figure 3. Results for Model no. 1. Comparison of the time evolution of the average number of bound sites, $\langle n \rangle$, for the matched target (crosses) and for the target with frustration at the fifth position (triangles). The target consists of $m = 10$ sites. The frustration factor is $f = 10$. The binding rate is $k_b^+ = 0.0001$. The equilibrium constant s is: A, 6; B, 1.6. Note the tenfold difference in the vertical scale. The lines represent a single exponential fit to the pseudo-first order kinetics.

whereas s depends on ambient conditions, mostly temperature and ionic strength. Figure 3 shows the results of Monte Carlo simulation of the kinetic behavior of this model for $k_b^+ = 10^{-4}$ and for two different values of s. These plots represent the average number of base-pairs, $\langle n \rangle$, formed between the 10 nt target and the 10 nt oligonucleotide as a function of time. The curves of each plot correspond to the correct target and the target with one mismatch in the middle.

One can see the dramatic difference in the behavior of our system for the two values of stability parameter. At high s ($s = 6$, Figure 3A), we observe gradual saturation of the target site with the asymptotic $\langle n \rangle$ close to its maximal possible value of 10. This indicates high affinity of binding. However, the curves for correct and mismatched binding are virtually identical indicating very poor specificity of binding. By contrast, for $s = 1.6$ (Figure 3B) $\langle n \rangle$ is much smaller than 10 even at saturation, which means very poor affinity of binding. At the same time the kinetic curves for correct and mismatched binding differ dramatically indicating very high specificity. The behavior of the system with respect to affinity and specificity is fully determined by the stability parameter. The binding rate k_b^+ only affects the pseudo-first-order kinetic constant of the reaction. Note that the first order kinetics perfectly describes all our data as is seen from the ideal single exponential fits shown in Figure 3.

The behavior of our system, i.e. high affinity at low specificity and *vice versa*, is in apparent contradiction with the conclusion of Eaton *et al*. (1995) described in the Introduction. To understand the reason for this contradiction let us analyze Model no. 1 within the framework of the vHB approach. In doing so we shall extend the standard derivation to address important features of Model no. 1. First, we are interested in considering not only the case when the target sites are poorly occupied, as is the case in Figure 3B, but also considering the case when occupancy is very high, like in Figure 3A. The latter case requires an accurate calculation of the $[T]$ value in equation (2) because $[T]$ is the concentration of free target sites. Second, the binding in Model no.1 occurs at many sites (residues) simultaneously. Only few of these sites are frustrated in a mismatched target. As a result, equilibrium constants K and $\hat{K}$ are not independent. If affinity to the matched target becomes too high, binding to the mismatched target could also be strong enough and specificity is compromised. This makes the oligonucleotide situation radically different from the case of ligand-protein and protein-protein recognition. Third, to facilitate the forthcoming analysis of the Model no. 2, we wish to investigate not only the equilibrium specificity but also the kinetic behavior of our system, as it is presented in Figure 3.

Let us assume a large excess of the oligonucleotide with respect to the target sites. We then can ignore variations of the oligonucleotide concentration and consider the following reaction:

$$[F] \underset{k_b^-}{\overset{k_b^+}{\rightleftharpoons}} [B] \tag{4}$$

where F stands for the free target and B refers collectively to all various bound microstates. We will discuss validity of such a collective description below. Let us denote fraction of free targets as F and the fraction of bound targets (occupancy) as B, so that $F + B = 1$. The kinetics of the reaction expressed by equation (4) obeys the single exponential law:

$$B(t) = B_0(1 - \exp(-t/\tau)) \tag{5}$$

where:

$$B_0 = \frac{k_b^+}{k_b^+ + k_b^-} = \frac{K_b}{1 + K_b} \tag{6}$$

is the equilibrium occupancy and $K_b = k_b^+/k_b^-$ is the effective equilibrium constant. The pseudo-first-order kinetic constant for the reaction is:

$$\tau^{-1} = k_b^+ + k_b^- \tag{7}$$

Note that the kinetic constant k_b^- is absent in the initial formulation of Model no. 1. Instead, we had constants k_e and k_r, which are interrelated *via* stability constant s. Quantity B_0 can be expressed in terms of k_b^+ and s. Thus equation (6) is actually a

definition of the effective decay constant k_b^-. The description of Model no. 1 by the simplified equation (4) is appropriate only when $k_b^- \ll k_e$. Then the bound macrostate ''lives'' sufficiently long for the equilibrium between all constituent microstates to be reached, and it can be adequately characterized by a single effective decay constant k_b^-.

Let us compare the situation for correct and mismatched binding. The value of k_b^+ is evidently mismatch-independent. By contrast, k_b^- is different for correct and mismatched targets. For sufficiently large s, it can be roughly estimated as (Anshelevich *et al.*, 1984):

$$k_b^- = s^{-m} \qquad \text{matched case} \tag{8a}$$

$$\hat{k}_b^- = s^{-m} \min(s^p, f) \qquad \text{mismatched case} \tag{8b}$$

Hereafter k_b^- is expressed in the units of k_e (we assume that $k_e = 1$). The value of p in equation (8b) depends on the position of the mismatch: it is the number of residues between the mismatched site and the nearest end of the target.

We can now consider how affinity and specificity of oligonucleotide binding depend on solution conditions. According to equation (8), k_b^- is a very strong function of s. When s is very close to unity (this corresponds to high temperature or low ionic strength) $k_b^- \sim 1$ for both correct and mismatched cases and the simplified collective description of the bound state is not applicable. However, neither affinity nor specificity can be large in this uninteresting case of very weak interaction between the ligand and its target. As s increases, both effective decay rates k_b^- and $\hat{k}_b^-$ decrease, but this decrease occurs much more slowly in the mismatched case than in the correct one. Using equation (8) and taking into account that k_b^+ is mismatched-independent, we find that the ratio of matched and mismatched equilibrium occupancies is:

$$\frac{B_0}{\hat{B}_0} = \frac{K_b + \min(s^p, f)}{K_b + 1} \tag{9}$$

where $K_b = k_b^+ s^m$. In the case of specific binding, $B_0/\hat{B}_0 \gg 1$, first term in the numerator of the above expression can be neglected. We see then that the specificity varies with s in two distinct patterns, depending on whether K_b is large or small as compared to 1 when $s^p = f$. For a single mismatch in the central part of the sequence, $p \approx m/2$, the behavior of the Model no. 1 depends on the value of the parameter $k_b^+ f^2$. When this parameter is small, $B_0/\hat{B}_0$ increases as s^p until it reaches maximal value f, stays at f until $k_b^+ s^m$ approaches 1 (significant affinity) and then decreases as s^m. In the opposite case, $k_b^+ f^2 \gg 1$, $B_0/\hat{B}_0$ increases as s^p reaching maximal value of approximately $(k_b^+)^{-1/2}$ at $k_b^+ s^m \approx 1$ and decreases. We conclude from this analysis that the specificity could be high only if binding rate k_b^+ is very low. The specificity increases with affinity in agreement with vHB only

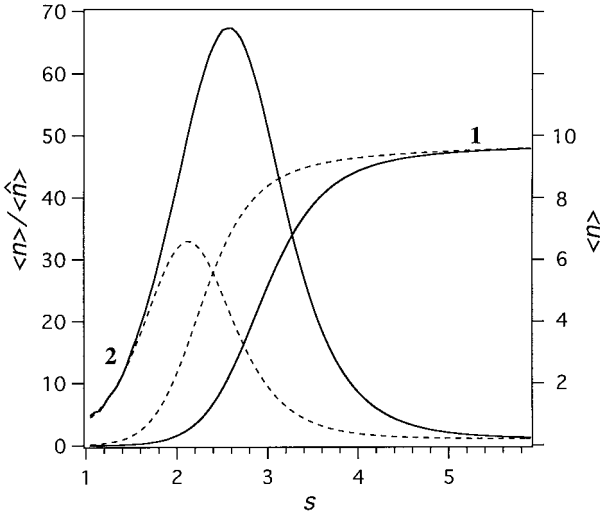

Figure 4. Results for Model no. 1. The affinity, $\langle n \rangle$, of the matched binding, curves 1 and the specificity, $\langle n \rangle / \langle \hat{n} \rangle$, curves 2 as functions of s. The mismatch is at the fifth position. Frustration factor $f = 160$. Broken curves, $k_b^+ = 0.0001$. Continuous curves, $k_b^+ = 0.00001$.

while the probability of matched binding remains small and starts to decrease rapidly as soon as B_0 approaches unity.

Figure 4 illustrates the above simple consideration. To facilitate the forthcoming analysis of Model no. 2, curves 1 in this Figure represent the average number of links formed between the target and the ligand, $\langle n \rangle = B_0 \bar{n}$. Here $\bar{n}$ is the average number of links in the bound state. Accordingly, we define the equilibrium specificity as the ratio $\langle n \rangle / \langle \hat{n} \rangle$ (curves 2). Figure 4 clearly shows an anticorrelation between affinity and specificity of binding of an oligonucleotide to its target site. Note that the location of the maximum on the affinity curve is determined by the condition $k_b^+ \sim s^{-m}$ and varies with k_b^+, which is proportional to the oligonucleotide concentration. However, upon changing the concentration, both curves in Figure 4 move along the s axis in a concerted way, so that their relative position is practically unchanged. Figure 4 also shows that the anticorrelation between affinity and specificity and the absence of the s value region where both affinity and specificity are high, remain well in place for a realistic frustration value $f = 160$.

We therefore conclude that in the case of binding of oligonucleotides to DNA or RNA *via* Watson–Crick pairing, affinity and specificity anticorrelate with one another. Note that most molecular biology techniques do not require the sequence recognition to be so stringent as we understand it throughout this work. The target sites and corresponding oligonucleotides are sufficiently long so that the probability to encounter sites with very few mismatches is negligible. One can expected to meet sites with numerous mismatches. As a result, there is a wide range of ambient conditions, where

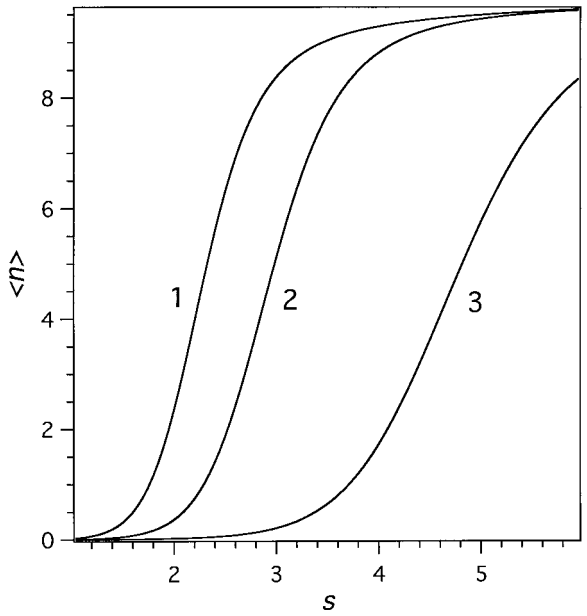

Figure 5. Results for Model no. 1. The affinity of binding to the ten-site target with: curve 1, no mismatches; curve 2, a mismatch at the fifth position; curve 3, three mismatches at second, fifth and seventh positions. The frustration factor is $f = 10$. The binding rate is $k_b^+ = 0.0001$.

$k_b^+ \gg k_b^-$ for the correct match while $\hat{k}_b^+ \ll \hat{k}_b^-$ for the (multiple) mismatches. An example is shown in Figure 5 where we see a wide range of the stability constant s with a high occupancy for the correct match and a low occupancy for the multiple mismatches. In this case, though affinity and specificity anticorrelate, there is still the possibility of both being sufficiently high.

Though in many practical applications the stringent sequence recognition of single-stranded nucleic acids by oligonucleotides is not necessary, cases definitely exist for which stringent specificity is essential. For such cases one should be aware of the anti-correlation effect between affinity and specificity that our computer simulations clearly demonstrate.

Specificity of recognition *via* Hoogsteen pairing

Duplex DNA can be targeted by triplex-forming oligonucleotides (TFO) *via* Hoogsteen pairing (reviewed by Frank-Kamenetskii & Mirkin, 1995; Soyfer & Potaman, 1996). Formally, the case is described by Model no. 1 just like the above case of Watson–Crick recognition. Our major conclusions therefore hold. Again, for long targets and long TFOs a range of ambient conditions exist where high affinity to correct target co-exists with low affinity to multiply mismatched targets. However, isolated mismatches are not associated with sufficiently large free energy loss so that affinity anti-correlates with specificity.

Specificity of PNA interaction with DNA

It was discussed in the Introduction that homopyrimidine PNAs present a special case, which is of great interest from the viewpoint of the specificity/affinity relation. These PNAs are capable of forming exceptionally stable complexes with complementary homopurine targets on single-stranded and even double-stranded DNA. An extraordinary affinity of homopyrimidine PNAs to their target sites on DNA is due to the unique properties of the PNA backbone. At the same time, the specificity of PNA–DNA interaction seems to be governed by essentially the same factors as in the case of the usual nucleic acids recognition, because specificity is determined by the same Watson–Crick and Hoogsteen base-pairing. Equations similar to equations (6) to (9) above would then inevitably lead to the conclusion that PNA should show very poor sequence specificity.

Surprisingly, experiments show that the exceptionally high affinity of homopyrimidine PNAs to target sites on double-stranded DNA is often accompanied by a remarkable specificity of interaction (Demidov *et al*., 1995, 1996; Veselkov *et al*., 1996a,b; Kuhn *et al*., 1998). This finding led Demidov *et al*. (1995; 1996) to the conclusion that the PNA-DNA complexes involve a new principle of biomolecular recognition. They claimed that affinity of PNAs to their DNA target sites is so high that the binding should be considered as irreversible. If this is true, equations (6) to (9) above, as well as the whole vHB-type treatment fails. However, although the overall process is irreversible, its first stage is highly reversible. According to Demidov *et al*. (1995), this reversible stage consists either of a fluctuation opening the DNA double helix and leading to the transient formation of the Watson–Crick duplex between the PNA oligomer and the complementary DNA strand (the ''Watson–Crick-first'' mechanism) or of the transient formation of the highly unstable $(DNA)_2$/PNA triplex *via* Hoogsteen pairing (the ''Hoogsteen-first'' mechanism). Only after the $(PNA)_2$/DNA triplex is formed by the association of the second PNA oligomer, does the binding become irreversible.

To check whether the formation of $(PNA)_2$/DNA triplexes can really entail a combination of high affinity and high specificity, we have performed extensive analysis of the behavior of the system described by Model no. 2 (Figure 2). Formally, the model considers the association of a single-stranded DNA target with two PNA oligomers. However, the model is equally applicable to the case of duplex target, if the complex is formed *via* strand invasion. The only difference is in the meaning of the parameter s. In case of the single-stranded target, the s has the same meaning as in case of Model no. 1, i.e. it is the stability constant for the Watson–Crick duplex elongation. Note, however, that the specificity of interaction of the

bisPNA and ssDNA has not been studied yet. In case of the duplex target, the meaning of the parameter s depends on the specific mechanism of the process. In the Watson–Crick-first mechanism, s is the ratio of the stability constant of the PNA/DNA duplex and the stability constant of the DNA/DNA duplex. In the Hoogsteen-first mechanism, s is the stability constant of the $(DNA)_2$/PNA triplex formation. Therefore, the results of computer calculations that follow are applicable, in principle, to both targets, single-stranded and double-stranded, and to both mechanisms, Watson–Crick-first and Hoogsteen-first. However, while discussing the results we will usually mean the duplex target because this reaction is better studied and is used in applications (Demidov *et al*., 1995, 1996; Veselkov *et al*., 1996a,b; Kuhn *et al*., 1997).

Model no. 2 corresponds to Model no. 1 with an additional step. We refer to this second step as the locking step. The two-step mechanism of binding leads to qualitatively different behavior of our system as compared to Model no. 1. In particular, within the framework of the two-step mechanism it becomes possible that the specificity of the overall binding reaction is determined by the first, binding step (i.e. is essentially the same as in Model no. 1), whereas the overall yield of the reaction of binding of PNA to DNA is determined by the second, locking step and thus can be completely independent of the specificity. Thus, in contrast to Model no. 1, Model no. 2 may provide both high affinity and high specificity of the overall binding reaction.

We have already approximated Model no. 1 by the first order kinetics assuming that all bound microstates are in fast quasiequilibrium with each other. Similarly, we may describe Model no. 2 by a two-step reaction:

$$[F] \underset{k_b^-}{\overset{k_b^+}{\rightleftharpoons}} [B] \underset{k_l^-}{\overset{k_l^+}{\rightleftharpoons}} [L] \tag{10}$$

Here, as in Model no. 1, F denote the fraction of unoccupied target sites, B stands for the fraction of target sites, which are occupied by one PNA oligomer (i.e. are bound but not locked), and finally, L is the fraction of the target sites which form triplex with two PNA oligomers (i.e. are locked). Clearly, $F + B + L = 1$. In reality, both macrostates, $[B]$ and $[L]$ are sets of many individual microstates. A simplified description given by equation (10) is applicable only when all rates in this equation are significantly slower than the rates of transitions between the various locked microstates and between the various bound unlocked microstates. In this case we may fully characterize locked and bound unlocked macrostates by the corresponding equilibrium distribution of microstates.

Let us first qualitatively identify relations between the various kinetic parameters of Model no. 2 that should be satisfied for high affinity and high specificity of the reaction. In our analysis of Model no. 1 we have seen that specificity of the binding is entirely due to the difference in the decay rates k_b^- of the state $[B]$ for matched and mismatched targets, and that maximal specificity requires low affinity, $k_b^+ \leqslant k_b^-$. From the transient state $[B]$ two routes are possible. First, the system can return back to the free state $[F]$ with the rate constant k_b^-. Secondly, it can proceed into the locked state with the rate k_l^+. To ''transmit'' the specificity of binding reached at the first stage to the final locking stage, the inequality $k_l^+ \ll k_b^-$ should be fulfilled. Otherwise, the constant k_b^-, which determines the specificity of the binding stage, will play no role in the overall kinetics. And last, to achieve high stability of the final complex, a strong inequality $k_l^+ > k_l^-$ should be assumed. With the above reasoning in mind, we henceforth shall focus our attention on the following range of kinetic parameters:

$$k_b^+ < k_b^-;\ k_b^- \gg k_l^+ > k_l^- \tag{11}$$

The two-stage reaction, equation (10), is described by the following kinetic equations:

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= -k_b^+ F + k_b^- B \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= +k_b^+ F - k_b^- B - k_l^+ B + k_l^- (1 - B - F) \end{aligned} \tag{12}$$

which can be easily solved. From equation (12) with $\dot{B} = \dot{F} = 0$ the equilibrium fractions of free (F), bound non-locked (B) and locked (L) states can be expressed as:

$$F = \frac{1 - B_0}{1 + B_0 K_l} \tag{13a}$$

$$B = \frac{B_0}{1 + B_0 K_l} \tag{13b}$$

$$L = \frac{B_0 K_l}{1 + B_0 K_l} \tag{13c}$$

where $K_l = k_l^+ / k_l^-$ is the equilibrium constant for the locking reaction and the equilibrium probability of binding in the absence of locking B_0 is defined by equation (6). In the range of kinetic parameters described by equation (11) the solution of equation (12) for the time evolution of the probability of any kind of binding $B(t) + L(t)$ with the initial condition $F(0) = 1$ is approximately given by:

$$L(t) + B(t) = L(1 - B_0)(1 - \mathrm{e}^{-t/\tau_0}) + B_0(1 - \mathrm{e}^{-t/\tau}) \tag{14}$$

where:

$$\tau_0^{-1} = k_l^+ B_0 / L \qquad \tau^{-1} = k_b^+ / B_0 \tag{15}$$

Note that according to equation (11) $\tau_0^{-1} \ll \tau^{-1}$.

We can now summarize the main features of Model no. 2. The kinetic behavior in this case can be approximated by two exponential processes: a

fast one with time constant τ and a slow one with time constant τ_0. The fast process describing the equilibrium with the intermediate non-locked bound state is the same as in Model no. 1 (see equations (5) to (7)). The slow process with the time constant $k_l^+ B_0/L$ describes the locking process. We are interested in the case of $B_0 \ll 1$, which corresponds to the low affinity of the first stage of the overall complex formation. In this case the slow process is responsible for a major part of the kinetic curve. Note also that the initial slope of the kinetic curve is independent of the rate of ''unlocking'', k_l^-.

The time evolution of the average number of links formed, both locked and unlocked, could be expressed in a form similar to equation (14):

$$n(t) = (n_f - \langle n \rangle)(1 - e^{-t/\tau_0}) + \langle n \rangle (1 - e^{-t/\tau}) \qquad (16)$$

where n_f is the equilibrium average fraction of linked or locked residues and $\langle n \rangle$ is an equilibrium fraction of residues that would be linked in the absence of locking, that is in the framework of Model no. 1. In the domain of kinetic parameters defined by equation (11) $\langle n \rangle \ll n_f \approx m$. This relation expresses the high affinity of the overall binding, whereas affinity of the first stage is small, so that the specificity of the first stage may still be high. In such a case $L \approx 1$, and the binding reaction is essentially of the first order with the pseudo-first-order rate constant $\tau_0^{-1} = k_l^+ k_b^+ / k_b^-$.

The above conclusions are confirmed by results of Monte Carlo simulations presented in Figure 6. This Figure shows the time evolution of the average number of linked bonds for the matched, A, and for the mismatched, B, chain (note the difference in the ordinate scales in the two cases).

The initial slope of the slow process is practically independent of k_l^- and therefore the kinetic specificity can be determined from the limiting case of the irreversible locking $s_l \to \infty$. In this case of infinite affinity of both matched and mismatched binding, $L \approx \hat{L} \approx 1$, the equilibrium specificity $R = L/\hat{L}$ is equal to 1 (no specificity at all). At the same time the kinetic specificity r (i.e. the ratio of time constants (or initial slopes) of the slow processes in matched and mismatched cases) is determined by the specificity of the first stage of the reaction:

$$r \equiv \frac{\hat{\tau}_0}{\tau_0} = \frac{k_1^+ B_0}{\hat{k}_1^+ \hat{B}_0} \qquad (17)$$

According to the above definition, the kinetic specificity is the ratio of the slowly growing occupancies of the matched and mismatched targets at times which are short as compared to characteristic time of locking process τ_0, but long as compared to characteristic time of binding τ. Note that this definition is meaningful only in the context of Model no. 2. On the other hand, in the most interesting case of truly irreversible locking, the notion of affinity of binding becomes meaningless (affinity becomes infinitely high). We therefore arrive at a very important conclusion that the sets of meaningful parameters describing the binding are different for the cases considered within the framework of Model no. 1 and within the framework of Model no. 2. In the former case one deals with traditional notions of equilibrium specificity and affinity of binding to matched and mismatched sites in accordance with von Hippel & Berg (1986). In the latter case these notions are replaced by notions of kinetic specificity and occupancy of matched and mismatched sites.

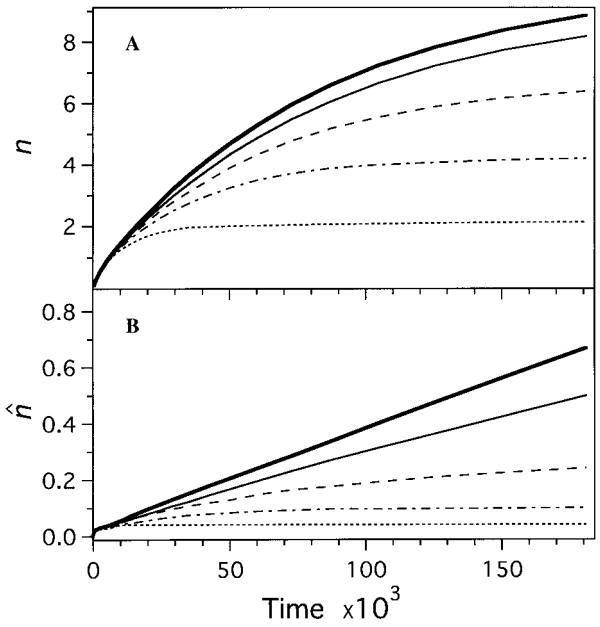

Figure 6. Results for Model no. 2. The time evolution of the average number of bound sites for the ten residue target for various locking stability constants, $s_l \equiv k_{le}/k_{lr}$. A; matched target, B; the mismatched target (at fifth position) for $f = 80$. The equilibrium constant for the binding is $s = 2.5$, the binding rate is $k_b^+ = 5 \cdot 10^{-6}$, the locking rate is $k_l^+ = 2 \cdot 10^{-5}$. From the bottom to the top: (·····) $s_l = 3$; (-·-·-) $s_l = 3.5$; (- - - -) $s_l = 4$; (———) $s_l = 5$; (▬▬) irreversible locking, $s_l = \infty$.

The locking rate constant k_l^+ generally varies for the different microstates within the bound macrostate B. One may assume that in the case of DNA/PNA complexes the locking rate is proportional to the number of links formed in a duplex. We have already argued that under conditions given by equation (11) the duplex state exists in quasi-equilibrium, and it is possible to ascribe one effective rate of locking for the whole bound macrostate. This locking rate is simply proportional to the average number of links in the bound state $\bar{n}$ (i.e. $k_l^+ = \bar{n}\bar{k}_l^+$, where $\bar{k}_l^+$ is the rate constant of locking of the individual link). Noting that $B_0\bar{n} = \langle n \rangle$, where $\langle n \rangle$ is the equilibrium fraction of links that would form in the absence of locking, we conclude that the kinetic specificity of the two-stage reaction is determined by the specificity of its first stage:

$$r = \langle n \rangle / \langle \hat{n} \rangle \qquad (18)$$

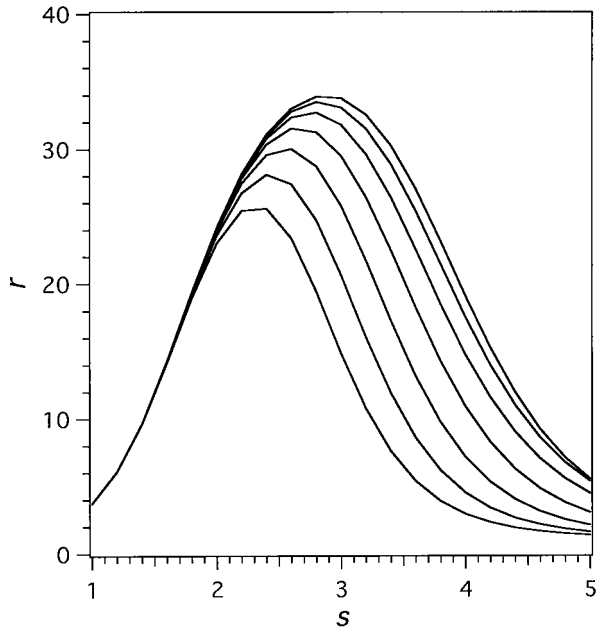

Figure 7. Results for Model no. 2. The kinetic specificity as a function of s. The parameter values are: $m = 10$; mismatch at the fifth position with $f = 40$; the binding rate is $k_b^+ = 5 \times 10^{-7}$. The locking rate decreases from curve to curve by a factor of 2 starting at $k_l^+ = 10^{-5}$ for the bottommost curve down to $k_l^+ = 1.563 \times 10^{-7}$ for the topmost curve.

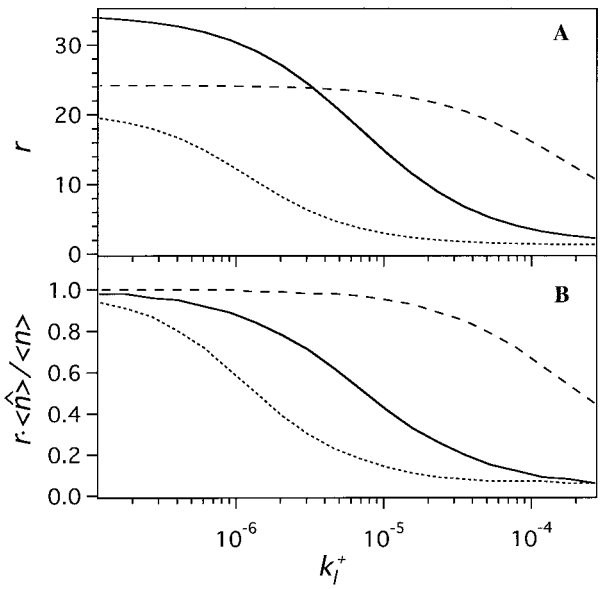

Figure 8. Results for Model no. 2. The kinetic specificity, r, A and the ratio of kinetic specificity to the equilibrium specificity B of the binding stage $r\langle\hat{n}\rangle/\langle n\rangle$ as a function of the locking rate, k_l^+. The parameter values are: $k_b^+ = 5 \cdot 10^{-7}$, $m = 10$, the mismatch at the fifth position, $f = 40$; dotted lines: $s = 2$; continuous lines: $s = 3$; broken lines $s = 4$.

The assumption that the locking is irreversible effectively means that the kinetics within the locked state is irrelevant. This greatly reduces the total number of states to be considered in the kinetic scheme and the problem becomes amenable to simple analysis by direct diagonalization of the matrix of kinetic coefficients. Figure 7 shows the kinetic specificity calculated as the ratio of the slowest relaxation times for the matched and mismatched binding in the case of irreversible locking. The data show that there is a relatively narrow region of the linking stability constant where specificity is high. Low specificity at small s is specific to our model. According to equation (8b) when $s^p \ll f$ and the advantage of forming new matched links is insufficient to compensate the formation of mismatched link, $\langle\hat{n}\rangle$ becomes independent of f and the specificity only slightly exceeds 1. Clearly, for $s < 1$ there should be no specificity. The specificity also decreases at large s. In part it is a manifestation of the general anticorrelation between affinity and specificity of the first stage of the reaction: as the stability of the complex formed at the first stage of the reaction increases its concentration saturates for both matched and mismatched targets. In addition, however, the dissociation rate of the bound state in the matched case can become small at large s as compared to the locking rate. The second of inequalities in equation (11) fails resulting in a decrease of the specificity relative to the prediction of equation (18).

This effect is illustrated in Figure 8. Here the kinetic specificity (A) and its relative deviation from the prediction of equation (18) are shown as a function of the locking rate for three s values. We see that the decrease in specificity at large k_l^+ parallels the deviation from equation (18). This shows unequivocally that the major cause of the reduction of the specificity in this case is a depletion of the stationary concentration of the intermediate bound state due to too fast locking. In other words, the specificity of recognition can be reached only if the final irreversible step of triplex formation is sufficiently slow so that quasi-equilibrium can be reached at the first stage of recognition. Otherwise, mismatched complexes would be ''locked'' due to the irreversible triplex formation with virtually the same rate as the correct complex.

An increase of the number of links in the chain dramatically increases the stability of the transient bound stage for the fixed s value. As a result, the specificity of the first stage and of the entire process drops much more steeply with increasing s. Figure 9 clearly demonstrates this trend. It shows the results of our calculations of the kinetic specificity as a function of s for targets containing 5, 7, 9, 11, 13 and 15 links with a mismatch at the central link. The data in Figure 9 predict that for each m there is the optimal value of s corresponding to the maximal kinetic specificity. The effect is more distinct for longer targets. This concrete theoretical prediction can be verified experimentally, since stability constant s can be varied greatly by changing ambient conditions. Moreover, because the

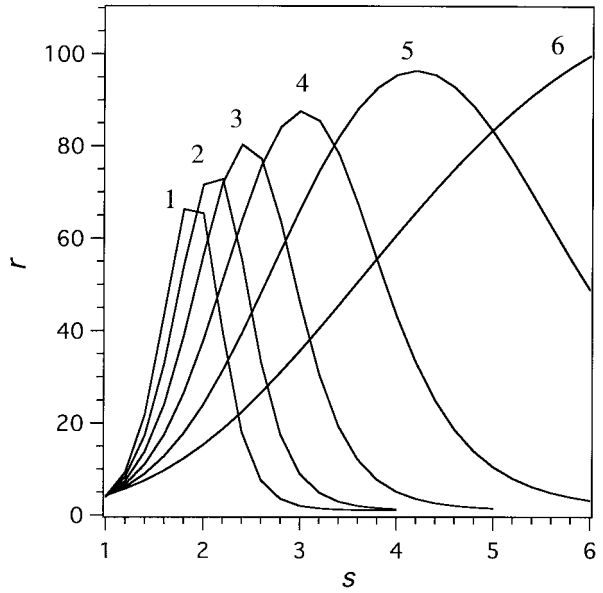

Figure 9. Results for Model no. 2. The kinetic specificity r as a function of s for various number of residues in the target site. The mismatch is located at the central position, $f = 160$, $k_b^+ = 5 \times 10^{-6}$, $k_l^+ = 10^{-7}$. (1) $m = 15$; (2) $m = 13$; (3) $m = 11$; (4) $m = 9$; (5) $m = 7$; (6) $m = 5$.

value of s has a completely different meaning for Watson–Crick-first and Hoogsteen-first mechanism, experimental study of curves as in Figure 9 may help to figure out what is the actual mechanism of the PNA/DNA binding. If the data in Figure 9 were confirmed experimentally, it would be a strong indication that our model of PNA/DNA interaction (Model no. 2) is correct. It would be also of paramount importance for the correct choice of the conditions corresponding to the maximal specificity of the complex formation.

Our analysis clearly demonstrates that the slow kinetics of triplex formation (a small value of k_l^+) is a precondition for the high specificity of the overall process. Available experimental data imply that bis-DNAs, like mono-PNAs, may exhibit high specificity of recognition (Veselkov *et al.*, 1996a,b; Kuhn *et al.*, 1998). This indicates that a high kinetic barrier for triplex formation is most probably an inherent property of PNA. This theoretical prediction needs to be tested by direct experiments. We may speculate, however, that the nature of the kinetic barrier depends on the specific mechanism of the process. It is easier to imagine this kinetic barrier in the case of the Hoogsteen-first mechanisms. Indeed, within the framework of this mechanism, the kinetic barrier may be associated with opening of the DNA duplex as the first step in the locking stage. The experimental data of Kuhn *et al.* (1998) indicate that the kinetic barrier for locking is sensitive to electrostatic interactions, because the kinetic specificity strongly depends on the distribution of positive charges along bis-PNA and on ionic strength.

It should be emphasized that the fact that the PNA interaction with duplex DNA exhibits kinetic rather than equilibrium specificity has serious ramifications for experimental procedures. The optimization of conditions for achieving significant occupancy of the correct target sites while maintaining minimal occupancy of mismatched sites has recently been theoretically analyzed in detail by Demidov *et al.* (1997). Kuhn *et al.* (1998) and Demidov *et al.* (unpublished results) have demonstrated for various PNAs the importance of the kinetic phenomena for the specificity of the PNA interaction with duplex DNA.

Above, we have concluded that PNAs with mixed composition do not have significant advantage over oligonucleotides in targeting single-stranded nucleic acids. What about homopyrimidine PNAs? Since Model no. 2 is valid in the case of interaction of the homopyrimidine PNAs with single-stranded nucleic acids, one can hope to observe both high specificity and high affinity for this case too. Paradoxically, the obstacle may consists in a too high stability of the PNA/DNA duplex. As we have shown, high affinity at the first stage of reaction seriously compromises the specificity of binding. We therefore predict a high specificity of targeting single-stranded nucleic acids under conditions corresponding to a relatively low stability of the PNA/DNA duplex. The data in Figure 9 are fully applicable to this case too, provided the parameter s has the meaning of the stability constant for the PNA/DNA duplex formation.

Conclusion

Within the framework of simple models, we have treated theoretically the problem of specificity of interaction of nucleic acids with oligonucleotides, their analogs and mimics. Standard assumptions, which are used in studying the specificity problem, are not applicable to these cases. In the case of oligonucleotides, the equilibrium stability constants for matched and mismatched binding change in a concerted way with changing ambient condition because only a small fraction of links is changed in the mismatch case. We have demonstrated that in the case of the interaction of nucleic acids with oligonucleotides specificity and affinity anticorrelate with one another for high values of the stability constant. In general, oligonucleotides and their analogs are imperfect candidates for truly sequence-specific recognition of nucleic acids. The lack of truly high specificity of oligonucleotides in targeting nucleic acids does not preclude their numerous applications because these applications actually do not require very high specificity of recognition.

Our theoretical results show that homopyrimidine PNAs, which bind to duplex DNA by forming a $(PNA)_2$/DNA triplex with one of the DNA strands, exhibit completely different behavior as compared to ordinary oligonucleotides and PNAs

with mixed purine-pyrimidine composition. Our calculations demonstrate that both high selectivity and high affinity can be reached in this case. Our theoretical data support the two-step mechanism of interaction of homopyrimidine PNAs with duplex DNA postulated by Demidov *et al*. (1995, 1996). The calculations show that a first reversible ''search'' stage followed by a second irreversible ''locking'' stage indeed allows the overall process to reach both high specificity and high affinity. We have also found the domain of parameter values that correspond to specific interaction between duplex DNA and homopyrimidine PNAs.

To describe a two-step binding process typical for complexes of PNA with dsDNA, we introduced the idea of kinetic specificity. The kinetic specificity is measured as a ratio of occupancy rates of matched and mismatched targets at times short as compared with the characteristic time of the locking process but long as compared with the characteristic time of the binding process. A notion of kinetic specificity is meaningless for the one-step binding reaction typical for oligonucleotide interactions. On the other hand, in the case of virtually irreversible locking, the notion of affinity of binding becomes meaningless (affinity is infinitely high). Instead, occupancies of matched and mismatched sites must be considered.

Although our predictions may not be quantitatively accurate due to the simplifications made, they give us qualitative understanding of the process. Further experimental studies should show whether or not our predictions are really met. Such a comparison of the theory and experiment will be a critical test of the proposed mechanism of interaction of homopyrimidine PNAs and dsDNA. Understanding of this mechanism is of paramount importance for design of further PNA-like drugs exhibiting high specificity and affinity of interaction with DNA.

The unique properties of homopyrimidine PNAs (and of bis-PNAs) have already been used to develop a new method for rare genome cutting (Veselkov *et al*., 1996a,b) and will undoubtedly find other applications in the field of nucleic acids. Our theoretical results confirm that in the case of homopyrimidine PNAs we encounter a new principle of biomolecular recognition.

Acknowledgements

We thank Drs Vadim Demidov, Peter Nielsen and Eugene Shakhnovich for fruitful discussions. This work was supported by NIH grant GM52201.

References

Anshelevich, V. V., Vologodskii, A. V., Lukashin, A. V. & Frank-Kamenetskii, M. D. (1984). Slow relaxational processes in melting of linear biopolymers: a theory and its application to nucleic acids. *Biopolymers,* **23**, 39–58.

Bond, J. P., Anderson, C. F. & Record, M. T. (1994). Conformational transitions of duplex and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients. *Biophys J.* **67**, 825–836.

Betts, L., Josey, J. A., Veal, J. M. & Jordan, S. R. (1995). A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex. *Science,* **270**, 1838–1841.

Cherny, D. Y., Belotserkovskii, B. P., Frank-Kamenetskii, M. D., Egholm, M., Buchardt, O., Berg, R. H. & Nielsen, P. E. (1993). DNA unwinding upon strand-displacement binding of a thymine-substituted polyamide to double-stranded DNA. *Proc. Natl Acad. Sci. USA,* **90**, 1667–1670.

Demidov, V. V., Yavnilovich, M. V., Belotserkovskii, B. P., Frank-Kamenetskii, M. D. & Nielsen, P. E. (1995). Kinetics and mechanism of peptide nucleic acid binding to duplex DNA. *Proc. Natl Acad. Sci. USA,* **92**, 2637–2641.

Demidov, V. V., Frank-Kamenetskii, M. D. & Nielsen, P. E. (1996). Complexes of duplex DNA with homopyrimidine peptide nucleic acid (PNA): new principle of biomolecular recognition. In *Biomolecular Structure and Dynamics* (Sarma, R. H. & Sarma, M. H., eds), vol. 2, pp. 129–134, Adenine Press, New York.

Demidov, V. V., Yavnilovich, M. V. & Frank-Kamenetskii, M. D. (1997). Kinetic analysis of specificity of duplex DNA targeting by homopyrimidine peptide nucleic acid. *Biophys. J.* **72**, 2763–2769.

Dueholm, K. L. & Nielsen, P. E. (1997). Chemistry, properties and applications of PNA (peptide nucleic acid). *New J. Chem.* **21**, 19–31.

Eaton, B. E., Gold, L. & Zichi, D. A. (1995). Let's get specific: the relationship between specificity and affinity. *Chem. Biol.* **2**, 633–638.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B. & Nielsen, P. E. (1993). PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen bonding rules. *Nature,* **365**, 566–568.

Egholm, M., Christensen, L., Dueholm, K. L., Buchardt, O., Coull, J. & Nielsen, P. E. (1995). Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-DNA. *Nucl. Acids Res.* **23**, 217–222.

Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995). Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95.

Frank-Kamenetskii, M. D., Anshelevich, V. V. & Lukashin, A. V. (1987). Polyelectrolyte model of DNA. *Sov. Phys. Usp.* **30**, 317–330.

Griffith, M. C., Risen, L. M., Grieg, M. J., Lesnik, E. A., Sprankle, K. G., Griffey, R. H., Kiely, J. S. & Freier, S. M. (1995). Single and bis peptide nucleic acids as triplexing agents: binding and stoichiometry. *J. Am. Chem. Soc.* **117**, 831–832.

Hammersley, I. M. & Handscamb, D. C. (1964). *Monte Carlo Methods,* Academic Press, London.

Hyrup, B. & Nielsen, P. E. (1996). Peptide nucleic acid (PNA): synthesis, properties and potential applications. *Bioorg. Med. Chem.* **4**, 5–23.

Kuhn, H., Demidov, V. V., Frank-Kamenetskii, M. D. & Nielsen, P. E. (1998). Kinetic sequence discrimination of cationic bis-PNAs upon targeting of

double-stranded DNA. *Nucl. Acid Res.* **26**, in the press.
Lukashin, A. V., Vologodskii, A. V., Frank-Kamenetskii, M. D. & Lyubchenko., Y. L. (1976). Fluctuational opening of the double helix as revealed by theoretical and experimental study of DNA interaction with formaldehyde. *J. Mol. Biol.* **108**, 665–682.
Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1991). Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science,* **254**, 1497–1500.
Nielsen, P. E., Egholm, M. & Buchardt, O. (1994). Evidence for $(PNA)_2$/DNA triplex structure upon binding of DNA to dsDNA by strand displacement. *J. Mol. Recogn.* **7**, 165–170.
Soyfer, V. N. & Potaman, V. V. (1996). *Triple-Helical Nucleic Acids,* Springer, New York.
Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science,* **263**, 77–784.
Tibanyenda, N., de Bruin, S. H., Haasnoot, C. A. G., van der Maler, G. A., van Boom, J. H. & Hilbers, C. W. (1984). The effect of single base-pair mismatches on the duplex stability of d(TATTAATATCAAGTTG) d(CAACTTGATATTAATA). *Eur. J. Biochem.* **139**, 19–27.
Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Norden, B. & Graslund, A. (1996). Ionic effects of the stability and conformation of peptide nucleic acid complexes. *J. Am. Chem. Soc.* **24**, 5544–5552.
Veselkov, A. G., Demidov, V. V., Frank-Kamenetskii, M. D. & Nielsen, P. E. (1996a). PNA as a rare genome-cutter. *Nature,* **379**, 214.
Veselkov, A. G., Demidov, V. V., Nielsen, P. E. & Frank-Kamenetskii, M. D. (1996b). A new class of genome rare cutters. *Nucl. Acids Res.* **24**, 2483–2488.
von Hippel, P. H. (1994). Protein–DNA recognition: new perspectives and underlying themes. *Science,* **263**, 769–770.
von Hippel, P. H. & Berg, O. G. (1986). On the specificity of DNA–protein interactions. *Proc. Natl Acad. Sci. USA,* **83**, 1608–1612.
Wittung, P., Nielsen, P. E., Buchardt, O., Egholm, M. & Norden, B. (1994). DNA-like double helix formed by peptide nucleic acid. *Nature,* **368**, 561–563.
Wittung, P., Nielsen, P. & Norden, B. (1996). Direct observation of strand invasion by peptide nucleic acid (PNA) into double-stranded DNA. *J. Am. Chem. Soc.* **118**, 7049–7054.

Edited by I. Tinoco

(Received 15 August 1997; received in revised form 20 October 1997; accepted 21 October 1997)