Dynamic, Thermodynamic, and Kinetic Basis for Recognition and Transformation of DNA by Human Immunodeficiency Virus Type 1 Integrase†

Dmitrii V. Bugreev,‡ Svetlana Baranova,‡ Olga D. Zakharova,‡ Vincent Parissi,§ Cécile Desjobert,§ Enzo Sottofattori,|| Alessandro Balbi,|| Simon Litvak,§ Laura Tarrago-Litvak,*,§ and Georgy A. Nevinsky*,‡

Novosibirsk Institute of Bioorganic Chemistry, Siberian Division of Russian Academy of Sciences, 8 Lavrentieva Avenue, Novosibirsk 630090, Russia, UMR 5097 CNRS-Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France and IFR 66 Bordeaux, France, and University of Genoa, Department of Pharmaceutical Sciences, Viale Benedetto XV, 3, Genoa-3, Italy

Received February 24, 2003; Revised Manuscript Received May 7, 2003

ABSTRACT: Specific interactions between retroviral integrase (IN) and long terminal repeats are required for insertion of viral DNA into the host genome. To characterize quantitatively the determinants of substrate specificity, we used a method based on a stepwise increase in ligand complexity. This allowed an estimation of the relative contributions of each nucleotide from oligonucleotides to the total affinity for IN. The interaction of HIV-1 integrase with specific (containing sequences from the LTR) or nonspecific oligonucleotides was analyzed using a thermodynamic model. Integrase interacted with oligonucleotides through a superposition of weak contacts with their bases, and more importantly, with the internucleotide phosphate groups. All these structural components contributed in a combined way to the free energy of binding with the major contribution made by the conserved 3′-terminal GT, and after its removal, by the CA dinucleotide. In contrast to nonspecific oligonucleotides that inhibited the reaction catalyzed by IN, specific oligonucleotides enhanced the activity, probably owing to the effect of sequence-specific ligands on the dynamic equilibrium between the oligomeric forms of IN. However, after preactivation of IN by incubation with Mn^{2+}, the specific oligonucleotides were also able to inhibit the processing reaction. We found that nonspecific interactions of IN with DNA provide ∼8 orders of magnitude in the affinity ($\Delta G° \approx -10.3$ kcal/mol), while the relative contribution of specific nucleotides of the substrate corresponds to ∼1.5 orders of magnitude ($\Delta G° \approx -2.0$ kcal/mol). Formation of the Michaelis complex between IN and specific DNA cannot by itself account for the major contribution of enzyme specificity, which lies in the k_{cat} term; the rate is increased by more than 5 orders of magnitude upon transition from nonspecific to specific oligonucleotides.

Replication of retroviruses depends on the integration of a double-stranded DNA copy of the retroviral genome into the host cell nuclear genome (reviewed in ref *1*). The integration step is catalyzed by the retroviral enzyme integrase (IN),[1] whose recognition sequence is located at the ends of the viral long terminal repeats. This LTR sequence is critical for site-specific cleavage and integration (*2*, *3*). HIV-1 integrase catalyses two reactions to insert both ends of the proviral DNA into the host cell genome: (i) 3′-processing, in which the two nucleotides (GT) from the 3′-ends of linear viral DNA are removed, leaving the CA dinucleotide at each 3′-end and (ii) the strand transfer or joining reaction in which the processed viral DNA ends are inserted into the host DNA (*4*, *5*).

Substrate recognition by IN is critical for retroviral integration (reviewed in ref *6* and references therein). To catalyze cleavage and strand transfer reactions, integrase must recognize viral DNA ends in a sequence-specific manner and host target DNA in a sequence-independent manner. The viral DNA sites for attachment to host DNA are at the termini and internal edges of the LTRs, yet only the terminal sites are used as attachment sites for integration. Studies on the interaction of IN with the DNA substrates have established that the viral enzyme can distinguish between viral DNA ends and other oligodeoxynucleotides. The most important sequence feature in specifying the viral attachment site is a CA/TG dinucleotide pair, invariably found precisely at the site of joining to host DNA. The presence of CA immediately upstream of the cleavage site in the LTR is a highly conserved feature of all retroviruses. These conserved bases

† This work was supported in part by grants from the Russian Foundation for Basic Research (01-04-49058), the Siberian Division of the Russian Academy of Sciences, the French Agence Nationale de Recherches sur le SIDA (ANRS), the Centre National de la Recherche Scientifique (CNRS), and the University Victor Segalen Bordeaux 2. V.P. benefited from a postdoctoral fellowship from SIDACTION. G.A.N. was supported by a short-fellowship from ANRS.

* Correspondence should be addressed to the following authors. (G.A.N.) Tel.: 7-3832 39 62 26. Fax: 7-3832 33 36 77. E-mail: nevinsky@niboch.nsc.ru. (L.T.-L.) Tel.: 33-5 5757 1740. Fax: 33-5 5757 1766. E-mail: laura.litvak@reger.u-bordeaux2.fr.

‡ Siberian Division of Russian Academy of Sciences.

§ UMR 5097 CNRS-Université Victor Segalen Bordeaux 2.

|| University of Genoa.

[1] Abbreviations: IN, integrase; HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; ON, oligonucleotide; ribo-ON, ribo-oligonucleotide; ODN, oligodeoxynucleotide; ds, double-stranded; ss, single-stranded; SILC, stepwise increase in ligand complexity.

10.1021/bi0300480 CCC: $25.00
Published on Web 07/02/2003

are crucial for recognition of viral DNA ends by IN. The site of 3′-end processing and DNA joining always corresponds to the 3′-OH of the conserved A. Positions just internal to the conserved CA also play a major role in susceptibility to integrase.

Viral DNA substrates must be double-stranded to be processed or DNA joined by integrase. In vitro, IN binds to substrate DNA with affinities similar to those of nonsubstrate DNA. The specificity of the reaction may involve the nonspecific binding of viral DNA first, followed by catalysis that is achieved in a complex way by the contribution of several nucleotides, both distal and proximal to the scissile bond, thus cleaving the CA in a site-specific fashion. Binding interactions extend inward at least 14−21 base pairs from the viral DNA end. There is no evidence that sequences further than 15 base pairs from the ends have any role in substrate specificity of IN. Therefore, most of the specificity appears to reside in the terminal base pairs.

In contrast to the specificity of integration for viral DNA ends, no obvious sequences have been defined for the large number of sites in host chromosomal DNA that can serve as the target for integration. A model has been proposed for the functional organization of IN in which viral DNA initially binds nonspecifically to the C-terminal portion of IN, and the catalytic central domain has an important role both in specific recognition of viral DNA ends and in positioning host DNA for nucleophilic attack (*7*). Target DNA recognition in vivo is probably influenced by structural features of the host DNA, as well as by the viral preintegration complex and cellular factors.

As protein−DNA interactions are involved in many of the fundamental processes that occur inside cells, it is extremely important to understand their nature. Many nucleic acid-binding proteins have low or no sequence specificity, whereas others have extremely high specificity for their special target sites. It is therefore particularly important to determine the molecular basis of specificity, which requires characterization of the conformational properties of the protein (and protein−ligand complex), the DNA target size, and the changes that ensue as a consequence of the interaction. Any differences observed in the specific complex need to be compared with any changes that occur in nonspecific complexes to be able to determine what constitutes specific binding (refs *8*−*10* and references therein).

In recent years, significant progress has been made in the detailed analysis of specific protein−DNA interaction using X-ray crystallographic techniques. However, X-ray structural analysis of protein−nucleic acid interactions does not provide quantitative estimates of the relative importance of molecular contacts, or of the relative contributions of strong and weak, or of specific and nonspecific contacts to the total affinity of an enzyme for DNA. X-ray analyses of sequence-specific enzymes with DNA have led to the somewhat erroneous concept that specific contacts, such as pseudo-Watson−Crick interactions, provide high affinity for specific DNA sequences and that such interactions lead to high specificity and high efficiency in catalysis. In addition, it is commonly supposed that enzymes recognizing specific double-stranded (ds) sequences cannot bind mononucleotides or short single-stranded (ss) oligonucleotides with high efficiency. Even though in vitro studies of specific binding use short DNA oligonucleotides, nonspecific binding can significantly contribute to binding equilibria for proteins whose specificity is not very large. Understanding the role DNA plays in facilitating the association of DNA-binding proteins is necessary for understanding how sequence specificity is accomplished. Therefore, the analysis concerning the quantitative evaluation of the relative individual contributions of the thermodynamic (complex formation) and kinetic (reaction rate constant) steps of a catalytic process to the DNA affinity or to the specificity of DNA-binding proteins is very important.

Different studies have been performed to address the energetic basis of specificity in DNA−protein complexes. Experiments using restriction endonucleases showed that *Bam*HI, *Eco*RI, and *Eco*RV have different context preferences, suggesting that the context affects binding by influencing the free energy levels of the complexes rather than that of free DNA (*11*−*13*). Jen-Jacobson described how enzymes that catalyze reactions at specific DNA sites have overcome the problem of inhibition by excess nonspecific binding sites on DNA (*14*). It has been proposed that the specific protein−DNA recognition complex bears a close resemblance to the transition state complex, such that very tight binding to the recognition site on the DNA substrate does not inhibit catalysis but instead provides energy that is efficiently utilized along the path to the transition state. More recently, the potential contribution of nonspecific binding to the thermodynamics of specific binding has been analyzed (*15*, *16*). Studies of the DNA-binding domains of several nuclear receptors reveal differences in structure and dynamics upon binding to DNA (*17*). As demonstrated recently, the *lac* repressor interacts with its natural operon through alternative conformations of its DNA-binding domain, thus showing a high degree of plasticity in DNA−protein recognition (*18*).

In general, site-specific protein−DNA complexes vary greatly in structural properties and in the thermodynamic strategy to achieve appropriate binding free energy. Overall, the primary and spatial structures of proteins, DNA, and RNA, as well as their conformational changes, possible plasticity, and additional protein−protein, RNA−RNA, and DNA−DNA interactions upon complex formation play an important role in the recognition and transformation of specific DNA and RNA sequences by enzymes and proteins (*9*, *10*). However, the factors governing the recognition and transformation of specific DNA may be different for each enzyme, thus making it difficult to find a relationship between them.

We have developed new approaches (*9*, *10*) to evaluate the relative contributions of individual nucleotides, including their structural elements, to enzyme affinity for DNA. The analysis of molecular interactions between enzymes and long nucleic acids by stepwise increase in ligand complexity (SILC) has shown that complex formation, including contacts between an enzyme and specific sequences, cannot alone provide the basis of either specificity or high affinity for DNA. Virtually all nucleotides within the DNA-binding cleft interact with the enzyme. High affinity (5−8 orders of magnitude) is mainly provided by numerous weak, additive (or close to additive) interactions between the enzyme and the various structural elements of nonspecific nucleotides. For specific DNA, the interaction of its nucleotides with enzymes may be additive or not and can include different

specific cooperative, anticooperative, or other interactions. However, in contrast to nonspecific DNA, the relative contribution of specific interactions to the total affinity is rather small, not exceeding 1−2 orders of magnitude. Thus, complex formation cannot alone explain the specificity of enzyme action. Specificity is provided by the enzyme-dependent DNA adaptation to the optimal conformation and by catalysis (*9*, *10*).

To better understand the basis for viral DNA-binding specificity of HIV-1 integrase, we used the SILC approach here. This method allowed us to probe interactions between integrase and a series of model substrates (single- or double-stranded, specific or nonspecific oligonucleotides) to characterize quantitatively the structural determinants of substrate specificity and the mechanism of action of integrase. Results were then analyzed using a thermodynamic model of specific DNA recognition.

EXPERIMENTAL PROCEDURES

Materials. Reagents were purchased from Merck and Sigma. Electrophoretically homogeneous HIV-1 integrase was purified from the JSC 310 protease-deficient yeast strain transformed with the integrase expression plasmid pHIV1SF2IN as previously described (*19*).

Oligonucleotides. Synthesis, purification, and characterization of homo- and hetero-oligonucleotides were performed as described before (*20*, *21*), and their concentration was determined as described by Fasman (*22*). The specific ds DNA substrate (termed ds GT-ODN$_{21}$) used for 3′-end processing was prepared by annealing the 21-mer ODN (5′-GTGTGGAAAATCTCTAGCAGT-3′) with the 21-mer complementary strand (5′-ACTGCTAGAGATTTTCCAC-AC-3′): both ODNs were heated for 2 min at 90 °C, followed by slow cooling. The ds ODN was then labeled at the 3′-end with [α-^{32}P]dGTP and [α-^{32}P]TTP in the presence of the exonuclease-free Klenow fragment of *Escherichia coli* DNA polymerase I (*19*, *23*).

Enzyme Activity Assay. Integrase activity was determined by measuring the 3′-end processing reaction at 30 °C. The standard reaction mixture (20−100 μL) contained 20 mM HEPES/NaOH (pH 7.5), 10 mM DTT, 0.1 mM EDTA, 4 mM NaCl, 7.5 mM $MnCl_2$, 0.05% Nonidet P-40, and 1.5−10 nM ds [^{32}P]GT-ODN$_{21}$ substrate. The reaction mixture was incubated for different times (2−60 min) in the presence of 5−40 nM IN. Products were quantified using two methods, both giving essentially identical results (*19*): (i) in the first method, the [^{32}P]GT product was separated by 15% PAGE in the presence of 7 M urea. Gels were autoradiographed, and gel pieces corresponding to the (^{32}P) cleavage products were measured by Cherenkov counting and (ii) in the second method, the reactions were stopped by transferring 10−50 μL aliquots to 40 μL of ice-cold solution of DNA (2 mg/mL) in 20 mM Tris-HCl containing 100 mM NaCl, and after thorough mixing, 1 mL of cold 8% trichloroacetic acid was added. Solutions were kept on ice for 3−4 h to allow precipitate formation. The precipitates were pelleted by centrifugation at 2 °C for 20 min (15 000 rpm), and 1 mL of the supernatants was used to count radioactivity. The content of [^{32}P]GT dinucleotides in the acid-soluble fraction was the same in both cases and contained <5% of the radioactivity typically observed after substrate incubation with IN. The following different controls were used: (i) the reaction mixture incubated with ds[^{32}P]-GT-ODN$_{21}$ in the absence of IN and (ii) the complete reaction mixture at zero time of incubation. All measurements were taken within the linear regions of time courses and enzyme concentration curves.

Effect of Oligonucleotides on the Rate of the 3′-Processing Reaction. To analyze the effect of oligonucleotides on IN activity, the enzyme was preincubated at 30 °C for 5−60 min in different conditions. The standard preincubation mixture (20−100 μL) contained 20 mM HEPES/NaOH (pH 7.5), 10 mM DTT, 0.1 mM EDTA, 50−100 mM NaCl, 2 mM CHAPS, 3% glycerol, 0.05% Nonidet P-40, 40−50 mM $MnCl_2$, 10 nM to 30 μM integrase, and different concentrations of oligonucleotides. At different time intervals, aliquots (5−10 μL) of the preincubated mixture were diluted with buffer (20 mM HEPES/NaOH, pH 7.5, 10 mM DTT, 0.1 mM EDTA, 4 mM NaCl) and added to the 3′-processing reaction mixture (final concentration of IN 5−10 nM), and the reaction was performed as described above.

Kinetic Parameters. Initial rates were measured in kinetic experiments. The K_M and V_{max} values were estimated using nonlinear regression analysis (*24*, *25*). The type of inhibition and the K_i values were determined by nonlinear regression analysis and presented as linear transformations using the Lineweaver−Burk plot. Values of IC_{50} were determined at a substrate concentration equal to $2K_M$. For competitive inhibition (*24*, *25*), $IC_{50} = 3K_i$ when $[S] = 2K_M$. Errors in IC_{50} and K_i were within 10−30%.

Small-Angle X-ray Scattering. The effect of oligonucleotides on the oligomerization of integrase was analyzed by preincubating the enzyme in 40 μL of a mixture containing 20 mM HEPES/NaOH pH 7.5, 10 mM DTT, 0.1 mM EDTA, 100 mM NaCl, 2 mM CHAPS, 3% glycerol, 0.05% Nonidet P-40, and 100 μM specific or nonspecific pentanucleotides. SAXS roentgenograms were obtained using a Siemens diffractometer (Germany) by the method of step-by-step scanning with a goniometer and X-ray scintillation detector as described earlier (*26*). The determination of the fraction composition using small-angle roentgenograms was performed as described previously (*26*).

RESULTS AND DISCUSSION

Complex Formation of HIV-1 Integrase with Oligonucleotides. The formation of the IN·DNA complex was analyzed using the SILC approach, according to the following scheme: orthophosphate or mononucleotide (as minimal ligands of IN) → ss nonspecific homo-d(N)$_n$ → ss specific hetero-d(N)$_n$ → ds nonspecific homo-d(N)$_n$ → ds specific hetero-d(N)$_n$.

We have previously shown that IN can bind different oligonucleotides: single- or double-stranded molecules of different lengths and with sequences, either related (specific) or unrelated (nonspecific) to the specific 21-mer substrate (*19*). Here, we extended these results to other oligonucleotides.

The 3′-end processing catalyzed by HIV-1 integrase was assayed with double-stranded GT-ODN$_{21}$, the specific DNA substrate derived from the HIV-1 U5 end of the LTR:

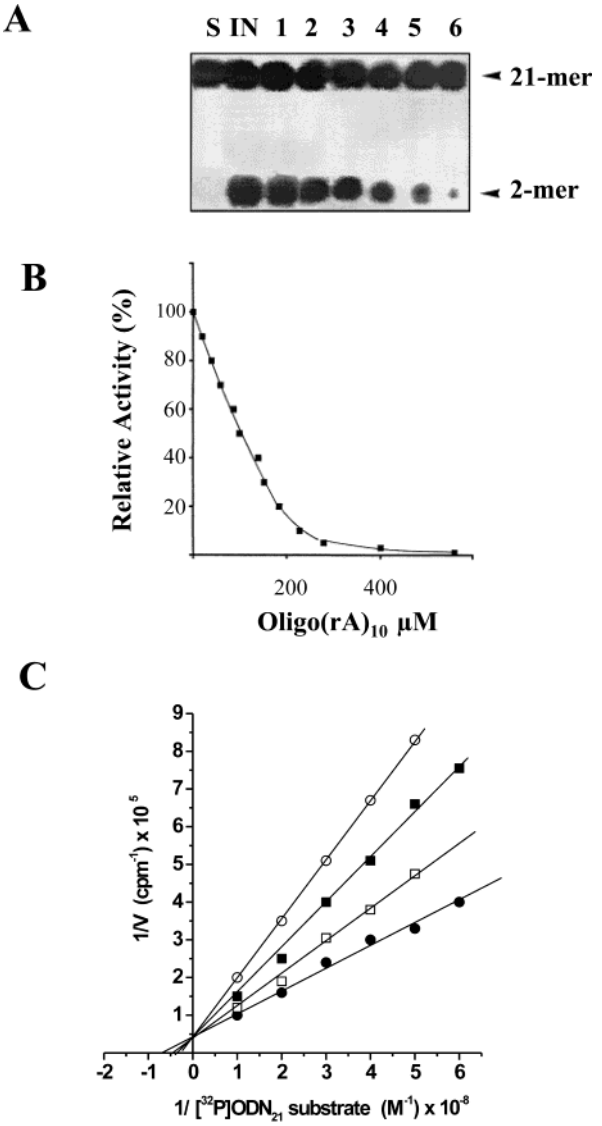

FIGURE 1: (A) Inhibition of the IN-catalyzed processing reaction by ribo-$(A)_{10}$. Integrase was incubated under the reaction conditions described in the Experimental Procedures, with the specific 21-mer duplex DNA substrate containing [^{32}P]-labeled GT nucleotides in the 3′-end. The reaction products were analyzed by 15% polyacrylamide−7 M urea gel electrophoresis. Complete reaction mixtures containing the labeled substrate were incubated in the absence of integrase (lane S) or in the presence of IN (lane IN). Lanes 1−6: same as lane IN but in the presence of 20, 50, 120, 150, 200, and 300 μM oligo $r(A)_{10}$. Arrows indicate the 21-mer and the dinucleotide product (2-mer). (B) Integrase activity. Reaction products were determined by measuring the TCA non-precipitable radioactivity corresponding to [^{32}P]GT remaining after IN-dependent processing. Results are expressed as relative activity, 100% corresponding to IN activity in the absence of inhibitor. (C) Double reciprocal plot according to the Lineweaver−Burk representation. The 3′-processing reaction was performed varying the GT-specific double-stranded [^{32}P]ODN_{21} substrate concentrations as indicated. Different concentrations of ribo-$(A)_{10}$ were added: 0 (●), 18 (□), 40 (■), and 88 μM (○). Values represent the average of at least two independent determinations. Errors were within 10−30%.

Kinetic analysis was performed by measuring the sequence-specific removal of [^{32}P]GT from the 3′-end (Figure 1A,B). A K_M of 1.3 ± 0.3 nM was determined for the specific [^{32}P]-GT-ODN_{21}. All nonspecific ONs inhibited the IN reaction competitively toward the substrate as illustrated for oligo-$(A)_{10}$ in Figure 1C. We showed that even short nonspecific ribo-ONs or orthophosphate competed with the ds GT-ODN_{21} substrate for the enzyme. The corresponding inhibition values for these oligonucleotides are reported in Table 1. Since substrate and inhibitors are competitive, the K_i value gives an estimate of the binding affinity ($K_d = K_i$) of IN for oligonucleotides. As most of the short oligonucleotides had relatively low affinities for IN, K_i values were calculated

Table 1: Affinity of HIV-1 Integrase for Nonspecific Oligonucleotides and Its Constituents[a]

(A) Nonspecific single-stranded oligonucleotides

$d(T)_n$	K_i (μM)	$d(C)_n$	K_i (μM)	$r(U)_n$	K_i (μM)	$r(C)_n$	K_i (μM)
T-base	>500000	C-base	>500000				
dTMP	767	dCMP	15000	UMP	4000	CMP	4000
T_2	330	C_2	300	U_2	2600	C_2	1500
T_3	166	C_3	100	U_3	2000	C_3	650
T_4	73	C_4	50				
T_5	43	C_5	32	U_5	**150**		
T_6	33			U_6	100	C_6	100
T_7	26	C_7	10				
T_{10}	**8.3**			U_{10}	30	C_{10}	40
		C_{16}	0.100				
T_{21}	**1.0**	C_{21}	0.012				

(B) Nonspecific single-stranded oligonucleotides and Pi[b]

$d(A)_n$	K_i (μM)	$r(A)_n$	K_i (μM)	$d(pR)_n(T)$[c]	K_i (μM)	compound[b]	K_i (μM)
A-base	>500000			D-ribose	>500000	Pi[b]	33000
dAMP	5000	AMP	7100	d(pR)	15000		
A_2	80	A_2	830				
A_3	48	A_3	310				
A_4	3.5	A_4	70	$d(pR)_3(T)$	100		
A_6	1.2	A_6	55				
A_8	1.0			$d(pR)_7(T)$	60		
A_{10}	**0.52**	A_{10}	**40**				
A_{12}	0.40						
A_{14}	0.30			$d(pR)_{13}(T)$	15		
		A_{18}	4				
A_{21}	**0.016**						
		A_{24}	2				

(C) Nonspecific double-stranded deoxy-oligonucleotides

no of units (n)	K_i: $d(T)_n$·$d(A)_n$ (μM)	ratio: K_i(ss)/K_i(ds) ss: $d(T)_n$ ds: $d(T)_n$·$d(A)_n$	ratio: K_i(ss)/K_i (ds) ss: $d(A)_n$ ds: $d(T)_n$·$d(A)_n$
3	45	3.6	1.06
4	3.2	23	1.1
6	0.600	55	2.0
8	0.280	57	3.6
10	**0.150**	55	3.4
12	0.078	70	5.1
14	0.054	83	5.6
21	**0.040**	25	4.0

(D) Nonspecific double-stranded deoxy- and ribo-oligonucleotides

ss ON	K_i (ss) (μM)	ds ON	K_i (ds) (μM)	ratio: K_i (ss)/K_i (ds)
dT_{10}	**8.3**	$d(T)_{10}$·$r(A)_{10}$	1	8.3
rU_{10}	30	$(rU)_{10}$·$d(A)_{10}$	0.26	115
rU_{10}	30	$(rU)_{10}$·$r(A)_{10}$	5	6
rA_{10}	**40**	$r(A)_{10}$·$d(T)_{10}$	1	40
dA_{10}	**0.52**	$d(A)_{10}$·$r(U)_{10}$	0.26	2
rA_{10}	**40**	$r(A)_{10}$·$r(U)_{10}$	5	8

[a] Values are the average results of two to three independent experiments. Errors in K_i and IC_{50} values were within 10−20%. K_i values (shown in bold) were determined by nonlinear regression analysis or calculated using the IC_{50} values (lightface) as described in the Experimental Procedures. [b] Pi corresponds to orthophosphate. [c] The $d(pR)_n$ series are deoxyribose phosphates in which R is a chemically stable analogue of deoxyribose (a tetrahydrofuran derivative).

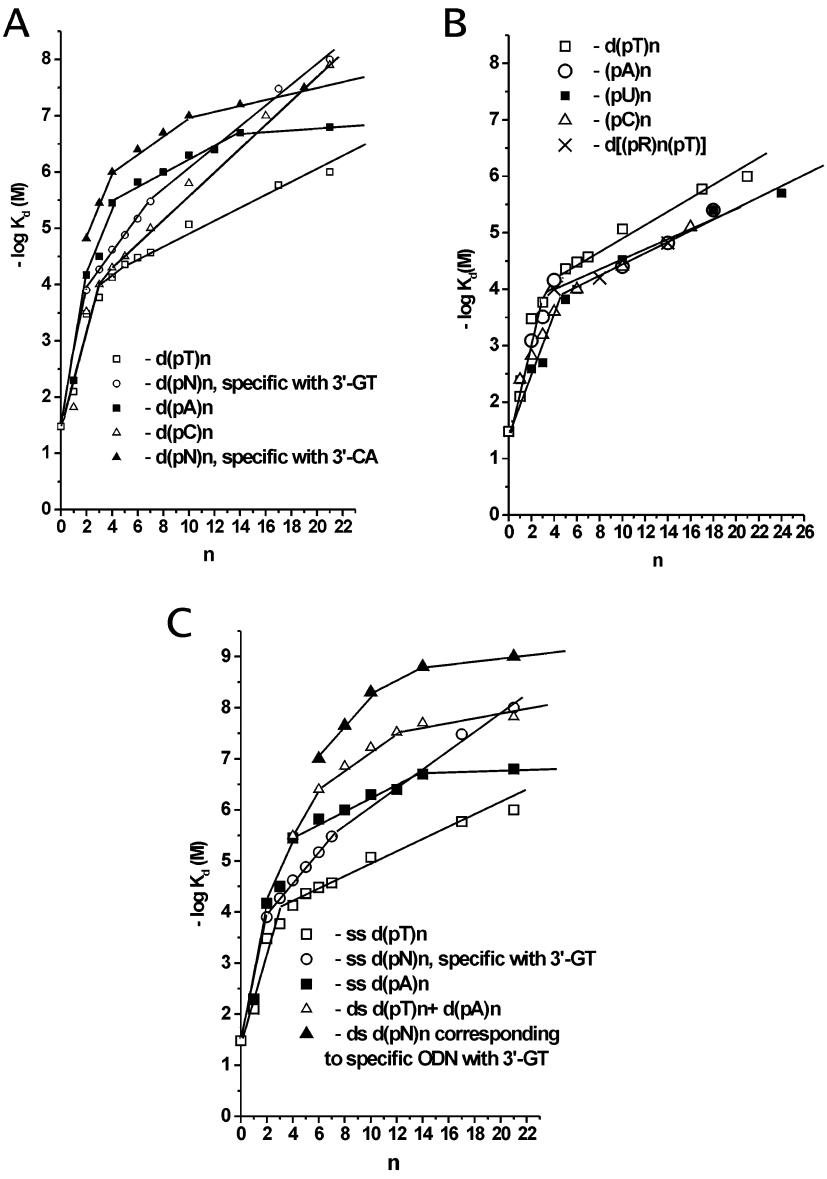

FIGURE 2: Dependencies of the logarithms of K_d (K_i) on the length of ss and ds oligonucleotides. Single-stranded deoxy-oligonucleotides (A); ss ribo- and deoxy-oligonucleotides (B); and ss and ds oligonucleotides (C). Symbols are shown in panels A−C. The −log K_d for orthophosphate corresponds to $n = 0$. Values represent the average of at least two independent determinations. Errors were within 10−30%.

from their IC_{50}. Under the conditions used, in which the concentration of the ds GT-ODN_{21} substrate was equal to $2K_M$, the IC_{50} corresponded to $3K_i$. This was confirmed experimentally with various ss and ds oligonucleotides. Thus, in most cases, the IC_{50} values were used to calculate K_i values (Table 1).

Interaction of HIV-1 Integrase with Nonspecific ss Oligonucleotide. Results in Table 1A,B show that while K_i values for Pi and dNMPs were comparable, the affinity for deoxyribose and for all bases was very low (K_i > 500 mM). We can thus conclude that even if IN recognizes free dNMPs through interactions with all its structural elements (base, sugar, and phosphate), the phosphate group makes the major contribution.

The Gibbs' free energy characterizing enzyme−ligand complex formation can be presented as the sum of $\Delta G°$ values for the individual contacts:

$$\Delta G° = \Delta G_1° + \Delta G_2° + ... + \Delta G_n°$$

with

$$\Delta G_i° = -RT \ln K_d^i$$

where K_d^i indicates the contribution of the individual contact (*24*, *25*). Hence, the overall K_d value characterizing complex formation is the product of the K_d values for individual contacts:

$$\Delta G° = -RT \ln K_d = -RT \ln[K_d(1)K_d(2) ... K_d(n)]$$

and

$$K_d = K_d(1)K_d(2) ... K_d(n)$$

To assess the possible additivity in the interactions of nonspecific oligonucleotides with IN, data from Table 1A,B were analyzed as the logarithmic dependencies of K_i (K_d) for $d(N)_n$ versus the number (n) of mononucleotide units ($0 \le n \le 21$, where $n = 0$ corresponds to Pi). As shown in

Figure 2A, all log-dependencies for nonspecific ss $d(N)_n$ appear biphasic. The transition from dTMP, dCMP, or dAMP to the corresponding $d(N)_2$ leads to an increase in the affinity by a factor of 2.5, 50, and 62, respectively. Values of *f*, defined as the increase in affinity of IN for various $d(N)_n$ per unit increase in length, were evaluated from the slopes of the linear parts of these curves. Integrase showed two regions close to linear at $n = 0-2$ and $n = 4-21$, respectively, while the region corresponding to $n = 2-4$ has a transitional character. Monotonic increases in K_d ($n > 4$), reflecting interaction between IN and each new nucleotide unit, are equal to the reciprocals of f ($K_i = 1/f = 0.79-0.33$ M). The K_i values characterizing the affinity of IN for dNMPs were higher (~1000 times) than the K_i characterizing the enzyme interaction with any other nucleotide of longer oligonucleotides ($n > 4$). The linear log-dependencies for ss oligonucleotides provide evidence of $\Delta G°$ values close to additive for the interaction of some of the 21 individual units of $d(N)_{21}$ with integrase.

We have previously shown that the interaction of different sequence-specific DNA enzymes (repair, topoisomerization, restriction enzymes) with each nucleotide unit of nonspecific ss or ds oligonucleotides is usually a superposition of weak electrostatic, hydrophobic, or van der Waals interactions with the individual structural elements (*10*). The interaction can be described by the descending geometrical progression:

$$K_d[(N)_n] = K_d[(P_i)](e)^{-n}(h_C)^{-l}(h_T)^{-m}(h_G)^{k}(h_A)^{-g}$$

where $K_d[(P_i)]$ is the K_d for the minimal orthophosphate ligand; *e* is a factor reflecting an increase of affinity because of one internucleotide phosphate group; and h_N are coefficients of increase in affinity because of hydrophobic and/or van der Waals interactions of the enzyme with one of the bases: C, T, G, and A, the numbers of which in $(N)_n$ are equal to *l*, *m*, *k*, and *g*, respectively. In addition, factor *f* is equal to (*he*). Only the values of *e* and h_N usually change from one enzyme to another. The affinity of some DNA-interacting enzymes for nonspecific $(N)_n$ does not always depend on the relative hydrophobicity of the bases. However, if the enzyme interacts with the bases, the increase in affinity for such oligonucleotides usually follows the same order as the increase in the relative hydrophobicity of the bases: C < T < G < A (9, *10*).

HIV-1 integrase was the first enzyme we found in which the affinity for nonspecific $d(C)_n$, $d(T)_n$, and $d(A)_n$ did not follow the relative hydrophobicity of the bases. Thus, the affinity of $d(A)_{21}$ and $d(C)_{21}$ (having minimal and maximal relative hydrophobicities) is similar, while the affinity of $d(T)_{21}$ is significantly lower (Table 1A,B). Interestingly, a faster increase in the affinity for $d(C)_n$ as compared with $d(T)_n$ resulted in a difference of ~2 orders of magnitude for the 21-mers (Figure 2A). As mentioned above, the interaction of some enzymes recognizing specific DNA can highly depend on the conformational flexibility and dynamic behavior of both enzyme and DNA and their capability for specific adaptation to each other. The higher affinity of $d(C)_n$ can also be explained by a better adaptation of $d(C)_n$ to a specific conformation in the viral DNA·IN complex and correlates well with a very flexible structure of $d(C)_n$, while $d(T)_n$ adopts a rigid nonflexible structure (*27*). To confirm the importance of the flexibility of $d(C)_n$ in its effective interaction with integrase, we compared the affinity of the enzyme for $d(T)_n$ and oligothymidylates containing several dC links in different positions (Table 2). The affinity for C/T-ODN_{21a} was comparable with that of $d(C)_n$ but not $d(T)_n$ of the same length (Table 1A). C/T-ODN_{21b} and C/A/T-ODN_9, having 4−5 T nucleotides in the 3′-end, presented an affinity intermediate between that for $d(T)_n$ and $d(C)_n$ of the same length. Therefore, the flexible structure of $d(C)_n$ is very important for a productive change during mutual adjustment of IN and DNA conformations.

Table 2: Affinity of Integrase for Nonspecific ss Oligonucleotides[a]

oligonucleotide	given name	No of units	K_i (μM)
CAT		3	20
AATT		4	140
GGAA		4	33
$(GA)_3$		6	1
$(AC)_3$		6	33
CCAACTTTT	C/A/T-ODN_9	9	8.3
(T_8)GT		10	8.3
$AC(A)_8$		10	18
TCACCTCCTT	C/T-ODN_{10}	10	**6.6**
CTGCGTCTATCAGCG		15	0.270
TTTTCCTCTCTCCCTCT	T/C-ODN_{17}	17	0.200
GGAAAATCTCTAGCAGT		17	0.033
GTGTGGAAAATCTCTAGCAGT	GT- specific ODN_{21}	21	0.010
CCCTCCTCCTTCTCTCTCCTT	C/T-ODN_{21a}	21	0.013
CCCTCCTCCTTCTCTCTTTTT	C/T-ODN_{21b}	21	0.100
CTAGCA**p**		6	2.6
$[Tp(Et)]_9T$[b]		10	8000
(T_9)p(ddT)		10	15
d(GAGATCGTC)**rA**		10	**0.500**

[a] Results are the average of two to three measurements. K_i values (shown in bold) were determined by nonlinear regression analysis or calculated using the IC_{50} values (lightface) as described in the Experimental Procedures. Errors in K_i and IC_{50} values were within 10−20%. [b] Ethylated analogue.

The contribution of the sugar moieties to binding was estimated by comparing the affinities for ribo- and deoxyribo-oligonucleotides. The structural characteristics of RNA and DNA differ markedly in solution: RNA usually exists in A form and DNA in B form. In addition, all $r(N)_n$ and $d(T)_n$ adopt very rigid nonflexible structures (*27*). The log-dependencies for all $r(N)_n$ were practically the same as those for $d(T)_n$ (Figure 2B). Therefore, for IN the superposition of enzyme interactions with each mononucleotide unit of ss $r(N)_n$ is essentially the same as for $d(T)_n$. Thus, since the affinity of different $d(T)_n$ and $r(N)_n$ does not in practice depend on their bases, it seems reasonable to suggest that IN interacts mostly with the sugar−phosphate backbone.

To probe the contribution of the sugar−phosphate backbone in the formation of many weak contacts between IN and oligonucleotides, we synthesized two different series of analogues: (i) the abasic oligomers, $d[(pR)_n(T)]$, in which R is a chemically stable analogue of deoxyribose with the base in C-1′ replaced by a hydrogen atom and (ii) the ethylated analogues. Table 1B shows that the affinity of the abasic analogues was slightly lower than the corresponding $d(T)_n$ (Figure 2B). The situation was different for the other analogues. The ethylation of the negatively charged internucleotide phosphate groups led to a decrease in affinity. For example, Table 2 shows that the $[Tp(Et)]_9T$ analogue had an affinity that is 3 orders of magnitude lower than that

of $d(T)_{10}$ (Table 1). These results further confirmed that integrase interacts through electrostatic interactions mainly with the phosphodiester backbone of oligonucleotides, a finding correlating well with data concerning the crystal structure of the HIV-1 integrase catalytic core and C-terminal domains (*28*). The structure resolved to 2.8 Å presents a Y-shaped dimer, in which the DNA binding site can be formed by the interaction of two monomers of the enzyme. The electrostatic potential map identifies a contiguous strip of positive charge along the outer face of the IN dimer, beginning at the active site of one monomer and expanding along the linker helix of the other monomer. This strip of positive potential may provide a base for IN interaction with internucleotide phosphate groups of specific and nonspecific DNA.

A change in affinity of 1.2−2.0-fold when elongating oligonucleotides ($n > 4$) by one nucleotide unit is lower (change in $\Delta G°$ of approximately −0.11 to −0.4 kcal/mol) than would be expected for strong electrostatic contacts (up to −1.0 kcal/mol) or hydrogen bonds (−2 to −6 kcal/mol) but comparable to values for weak ion−dipole and dipole−dipole interactions (*24*). Thus, the interaction of negatively charged internucleotide groups of nonspecific oligonucleotides with the DNA-binding channel of IN might rely on dipolar electrostatic interactions rather than on electrostatic interactions of immediate contacting groups. Since the structures of ribo-ONs, $d(T)_n$, and $d[(pR)_n(T)]$ cannot easily be adjusted by the enzyme, a decreased affinity of IN for the 5′-flanks of these ligands may result from the longer distance between the positively charged surface of IN and the sugar−phosphate backbone of these oligonucleotides. The increased efficiency of the interaction between IN and the 5′-flanks of ss $d(C)_n$ or specific ODNs could be due to bringing the oppositely charged surfaces of IN and oligonucleotides closer together as a result of easier conformational changes in DNA and the enzyme structure.

Affinity of IN for Nonspecific DNA Duplexes. Usually enzymes interacting with ds oligonucleotides contact both chains of DNA after partially melting them, thus maintaining the base-pairing between both strands. However, the contribution of the second strand to the affinity is usually much lower than that of the first strand (*10*).

To assess the contribution of the second strand, we analyzed the interaction of IN with nonspecific $d(T)_n \cdot d(A)_n$ duplexes. The affinity increased with the length of the duplex to a maximal value at $n = 12-21$ (Table 1C). The minimal ligand exhibiting duplex properties toward IN was the mixture of complementary hexanucleotides, $d(T)_6$ and $d(A)_6$, for which the T_m in solution (*29*) is significantly lower (<5 °C) than the reaction temperature (30 °C). Thus, for short duplexes, some stabilization is obtained by interaction with IN.

Next, we assayed ribo-oligonucleotides (Table 1D). Changing from $d(T)_{10} \cdot d(A)_{10}$ to $r(U)_{10} \cdot r(A)_{10}$ led to a decrease in affinity by a factor of ∼30, while in the case of mixed ribo-deoxy duplexes, $r(U)_{10} \cdot d(A)_{10}$ and $r(A)_{10} \cdot d(T)_{10}$, the changes in affinity were much lower. It is very likely that IN cannot efficiently force a productive conformation on the structure of an RNA−RNA. The replacement of one ribo- by a deoxyribo-strand led to an easier change in conformation, which was dependent on IN. Ratios between the K_i values of ss and ds oligonucleotides were highly dependent on the length and structural characteristics of these compounds (Table 1C,D). All these results show that integrase interacts with both strands of the DNA substrate but that the contribution of the second strand is significantly lower than that of the first. Thus, the rules governing the recognition of nonspecific ss and ds oligonucleotides by IN are very similar.

Activation of IN by Specific ODNs. As previously shown, the GT dinucleotide produces an inhibitory effect on IN activity similar to that observed with nonspecific oligonucleotides (*19*). However, an unexpected result was obtained with longer ODNs: the activity of the enzyme was stimulated after preincubation of IN with low concentrations of specific ODNs (Figure 3A). The level of activation increased with the length of ODNs (lanes 2−12). Specific ODNs containing CA in the 3′-end were better activators (lanes 7−12) than GT containing ODNs of the same length (lanes 2−6). The stimulation was even higher in the presence of ds ODNs corresponding to the 3′-terminal sequence of ODN_{21} (compare lanes 15 with 3 and 16 with 5). On the contrary, a nonspecific GT-containing compound such as $(T)_8(GT)$ did not activate the enzyme, whereas preincubation with CA or $(T)_8(CA)$ led to a small but detectable increase in the activity (lanes 7 and 8). The effect of ODNs corresponding to the noncleavable strand of DNA was notably lower (lanes 13 and 14). Lane 17 shows the activation of the enzyme after preincubation with Mn^{2+} ions, an effect previously described by us and others (*19, 30, 31*).

We then analyzed the relative activity of IN in the 3′-processing reaction as a function of the substrate concentration. This reaction usually requires a long incubation time (0.5−1 h), thus suggesting that even without preincubation, specific ODNs can to some extent activate IN during the course of the reaction. To check this, we analyzed the activity of IN as a function of the substrate using a fixed concentration (2 nM) of ds $[^{32}P]$GT-ODN_{21} and a stepwise increase in its total concentration (up to 62 nM) because of addition of nonradioactive ds ODN_{21} (Figure 4). Such isotopic dilution of a radioactive substrate should normally cause a decrease in the accumulation of the labeled product. However, the increase in total concentration of ODN_{21} led to a significant enhancement of $[^{32}P]$GT dinucleotide removal. The maximal level of apparent substrate-dependent activation of IN was about 6-fold. This was significantly lower than the real level of IN activation since adding 60 nM nonradioactive ODN_{21} to 2 nM $[^{32}P]ODN_{21}$ diluted the specific activity of the substrate by a factor of 31. Thus, the level of IN activation was so high that the activated enzyme molecules were able to remove GT efficiently, not only from nonlabeled ODN_{21} but also from $[^{32}P]ODN_{21}$.

The concentration of the ODN_{21} substrate (1−2 nM) in the reaction mixture is usually ∼10−100 times lower than that of the enzyme (10−100 nM). In general, the catalytic activity of HIV-1 IN is so low that most activity-based assays require more than stoichiometric amounts of the enzyme as compared to the concentration of the substrate (*31*). At fixed low concentrations of the ODN_{21} substrate, an increase in enzyme concentration up to ∼1 μM leads to a linear increase in the 3′ processing reaction (data not shown). This means that <1−5% of IN molecules are active in the reaction. It has been reported that IN can exist in a dynamic equilibrium of different oligomers. In the micromolar range, IN form high-order multimers such as tetramers, octamers, or ag-

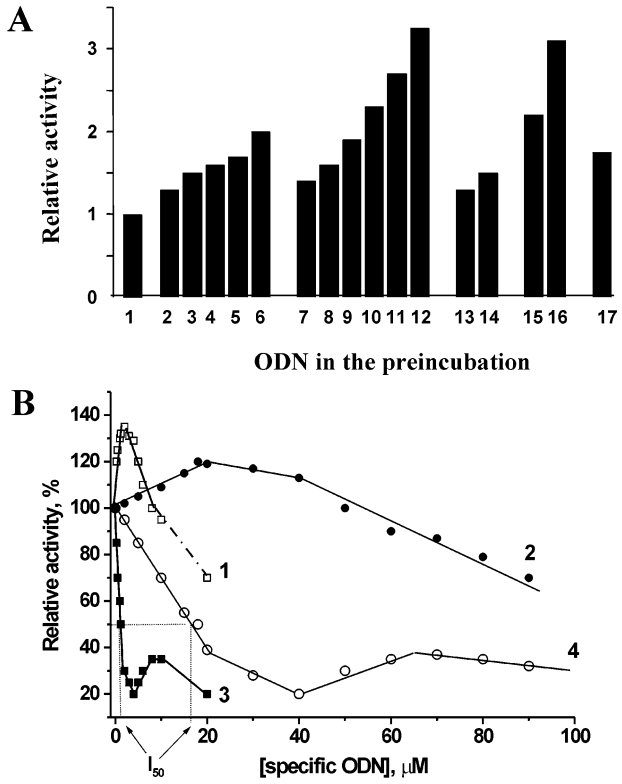

FIGURE 3: (A) Relative activity of IN in the processing reaction. The activity was measured using 10 nM GT-specific ds [^{32}P]ODN$_{21}$ as substrate. Integrase (2 μM) was previously preincubated for 1 h at 30 °C under different conditions: 1, in the absence of ODNs or $MnCl_2$ (this relative activity was taken as 1); 2−16, in the absence of $MnCl_2$ but after addition of 2.5 μM ODNs as follows: 2, AGT; 3, ss GT-specific hexamer (5′ AGCAGT 3′); 4, GT-specific ss octamer (5′ CTAGCAGT 3′); 5, GT-specific ss decamer (5′ CTCTAGCAGT 3′); 6, GT-specific ss ODN$_{21}$ (5′ GTGTGGAAAATCTCTAGCAGT 3′); 7, CA dinucleotide; 8, ss $(T)_8(CA)$; 9, ss CA tetramer (5′ AGCA 3′); 10, ss CA hexamer (5′ CTAGCA 3′); 11, ss CA decamer (5′ ATCTCTAGCA 3′); 12, ss CA-ODN$_{19}$ (5′ GTGTGGAAAATCTCTAGCA 3′); 13, a noncleavable decamer (5′ AAAAGGTGTG 3′); 14, a noncleavable 19-mer (5′ ACGATCTCTAAAAGGTGTG 3′); 15, double-stranded GT-specific hexamer (5′ AGCAGT 3′$(N)_6$); 16, double-stranded GT-specific decamer (5′CTCTAGCAGT 3′$(N)_{10}$); and 17, 40 mM $MnCl_2$ in the absence of ODN. (B) Effect of specific ODNs on the rate of 3′-processing. The reaction mixture was incubated for 30 min at 30 °C in the presence of 2 nM GT-specific ds [^{32}P]ODN$_{21}$ substrate and different concentrations of CTCTAGCA (lines 1 and 3) or CTAGCAGT (lines 2 and 4). The reaction was started by the addition of 10 nM integrase before (lines 1 and 2) or after preincubation with 40 mM $MnCl_2$ (lines 3 and 4). The relative amount of [^{32}P]GT in the acid-soluble fraction for reaction mixtures in the absence of oligonucleotides was taken as 100%. Values represent the average of at least two independent determinations. Errors were within 10−30%.

gregates (*32*). However, at catalytically active concentrations (10−100 nM), integrase mostly exists as monomers that are catalytically inactive. Therefore, the activation of IN after preincubation with specific ODNs might result from a change in the equilibrium between monomeric and oligomeric forms in favor of the formation of catalytically active oligomers. The low affinity of IN subunits for each other and the increase in the activation effect by specific ss and ds ODNs with increasing concentrations of IN strongly suggest that the preincubation of IN with such ODNs leads to the formation of catalytically active oligomeric forms of the enzyme.

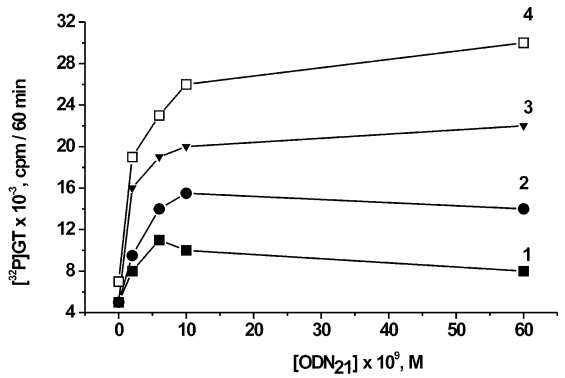

FIGURE 4: Processing activity as a function of oligonucleotide and IN concentrations. The activity was determined as a function of the concentration of nonradioactive ds GT-ODN$_{21}$ at a constant concentration of ds 3′-[^{32}P]GT-ODN$_{21}$ (2 nM) specific substrate. Integrase was used at different concentrations: 50 (1), 128 (2), 320 (3), and 700 nM (4). The relative accumulation of [^{32}P]GT was compared after 1 h of reaction. Before adding IN to the reaction mixture, the enzyme (2 μM) was preactivated by incubation with 40 mM $MnCl_2$ at 30 °C for 1 h. Values represent the average of at least two independent determinations. Errors were within 10−30%.

The relative amounts of different enzyme oligomers can be estimated using small-angle X-ray scattering (SAXS) (*26*, *33*). According to our preliminary data on IN analysis by SAXS, more than 95% of IN (100 nM) exists in monomeric form, while preincubation (for 0.5−1 h at 30 °C) with a specific GCAGT oligonucleotide leads to the formation of enzyme dimers. As was shown above, nonspecific ODNs compete with the substrate for IN, but under the preincubation conditions, nonspecific $d(T)_5$ or $d(A)_5$ did not lead to formation of IN dimeric forms. This indicates that nonspecific oligonucleotides interact predominantly with the preformed catalytically active oligomeric forms of the enzyme.

Besides several factors governing specificity, the viral DNA sequence-dependent formation of catalytically active oligomers of IN can be an important way to increase enzyme specificity. The propensity of IN for self-assembly increases with the length of specific GT-ODNs, but CA-containing ODNs corresponding to the cleaved strand are better activators. Therefore, it cannot be excluded that, in addition to their effect on IN self-assembly, these GT- and CA-containing ODNs can change the oligomer conformation by alternative routes. Our results on sequence-dependent activation indicate that IN is an extremely dynamic enzyme in conformational terms. According to the crystal structure, some residues in the active site region and between the α5 and α6 helices of catalytic core dimers form very flexible loops (*28*). It is possible that easily achievable changes in the structure of IN are necessary to provide different types of interactions with nonspecific and specific DNAs before and after the removal of the GT dinucleotide and the following reaction of integration. These results probably indicate sequence-dependent formation of catalytically active oligomers of IN.

Affinity of Integrase for Specific Oligonucleotides. The addition of low concentrations of specific ODNs to the reaction mixture led to activation of IN, whereas increasing their concentration inhibited the reaction (Figure 3B). Thus, the activation of IN made it difficult to estimate the affinity of specific ODNs for the enzyme when using the method of inhibitory analysis.

Table 3: Affinity of Integrase for Specific ODNs[a]

A

3'-GT-specific ODNs	N*	K_i (μM)	3'-CA-specific ODNs	N*	K_i (μM)
Pi	0	33 000			
dGMP	1	15 000			
GT	2	**130**	CA	2	**15**
AGT	3	53.3	GCA	3	3.5
CAGT	4	23.3	AGCA	4	1
GCAGT	5	**13**	TAGCA	5	0.700
AGCAGT	6	6.1	CTAGCA	6	**0.400**
TAGCAGT	7	3.3	TCTAGCA	7	0.300
CTAGCAGT	8	2.5	CTCTAGCA	8	0.200
CTCTAGCAGT	10	**0.500**	ATCTCTAGCA	10	0.100
AAATCTCTAGCAGT	14	0.180			
			GGAAAATCTCTAGCA	15	0.075
GAAAATCTCTAGCAGT	16	0.063			
GTGTGGAAAATCTCTAGCA	19	0.030	GTGTGGAAAATCTCTAGCA	19	0.030
GTGTGGAAAATCTCTAGCAG	20	0.150			
GTGTGGAAAATCTCTAGCAGT	21	0.010			
T_{19}GT	21	0.400			
TTTTGTAAAACCACGGCCAGT	21	0.170			
GTGTGGAAAATCTCTAGCAG**r**U	21	0.083			

B

Specific double stranded ODNs	K_i (μM)	Ratio : Ki ss/Ki ds
AGT•$d(N)_3$	40	1.3
CAGT•$d(N)_4$	10	2.3
AGCAGT•$d(N)_6$	0.1800	33.9
CTAGCAGT•$d(N)_8$	**0.0320**	78
CT CTAGCAGT•$d(N)_{10}$	**0.0100**	50
AAATCTCTAGCAGT•$d(N)_{14}$	0.0032	56.3
GAAAATCTCTAGCAGT•$d(N)_{16}$	0.0020	31.5
GTGTGGAAAATCTCTAGCA•$d(N)_{19}$	0.0040	7.5
GTGTGGAAAATCTCTAGCAG•$d(N)_{20}$	0.0600	2.5
GTGTGGAAAATCTCTAGCAGT•$d(N)_{21}$	0.0013	7.7
T_{19}GT•A_{19}CA	0.0360	11.1
GTGTGGAAAATCTCTAGCAG**r**U•$d(N)_{21}$	0.0500	1.7

[a] (A) Single-stranded ODNs. **N* corresponds to the number of nucleotide units. (B) Double-stranded ODNs. $d(N)_n$ corresponds in each case to the complementary strand. Results are the average of three to four measurements. The K_i values were determined by nonlinear regression analysis (in bold) or calculated using the IC_{50} values. The ratio K_i (ss)/K_i (ds) was calculated with the K_i values (ss) given in part A.

To evaluate inhibition, it was important to eliminate the activation produced by specific ODNs on IN (Figure 3B, lines 1 and 2). As shown above, preincubation of IN with Mn^{2+} increased the specific activity of the enzyme. We therefore preincubated IN in the presence of Mn^{2+} to obtain the activated form of the enzyme, and the processing reaction was then performed in the presence of specific ODNs. Figure 3B presents the results with two specific ODNs. When starting with a Mn^{2+}-activated form of IN, an inhibition similar to that obtained with nonspecific ODNs was obtained (Figure 3B, lines 3 and 4). In these conditions, specific ODNs were also found to be competitive inhibitors toward the substrate. The K_i values determined for specific ODNs are summarized in Table 3.

Complex Formation of IN with Specific ODNs: Interactions with the Nucleotide Units. As shown in Figure 2, all log-dependencies for specific and nonspecific ss oligonucleotides are very similar and appear nearly biphasic. The transition from dTMP, dGMP, or dCMP to the corresponding $d(N)_2$ or d(GT) increased their affinity by a factor of 2−115, while further lengthening by one nucleotide results in a nearly monotonic increase by a factor of ∼1.26−1.64. In contrast, the transition from dAMP to d(AA) or d(CA) and then to the corresponding $d(N)_4$ led to a significantly higher increase in affinity (4−330-fold). The increase became slow and nearly monotonic ($f = \sim 1.2$) only at $n = 5-21$. Thus, IN shows two different modes of interaction with specific short ODNs containing either the 3′-terminal GT (and related nonspecific $d(T)_n$ and $d(C)_n$ oligonucleotides) or the 3′-terminal CA (and related $d(A)_n$).

Before the processing reaction, the contribution of GT terminal nucleotides of ODN_{21} to the total affinity was significantly higher than that of any other nucleotide. However, the high affinity of CA-containing ODNs and their great ability to activate the enzyme suggest that after IN-dependent removal of the GT-terminal nucleotide, a change in the enzyme−DNA complex occurs. IN forms new strong contacts with the four AGCA-terminal nucleotides and probably retains to some extent the weak contacts previously formed with the 5′-flank of ODN_{21}.

In Table 3 it is shown that the K_i of AGCA is 1 μM. On the other hand, the contribution of the AGCA sequence within ss GT-ODN_{21} to the affinity of IN could be approximately estimated from the ratio of K_i values for

AGCAGT (6.1 μM) and GT (130 μM) (Table 3). This calculation gave a K_i of 0.047 M, showing that the contribution of the AGCA tetranucleotide to the enzyme affinity for CA-ODN_{19} was 4.7×10^4 times greater that that for GT-ODN_{21}. Moreover, from the ratio of K_i values for CA (15 μM) and AGCA (1 μM), it can be calculated that the contribution of the CA-terminal dinucleotide to the affinity of CA-ODN_{19} is about 5.5×10^4 times higher than that for the two additional AG nucleotides (0.067 M) of the AGCA terminal tetranucleotide.

All these data indicate that the 3′-terminal GT dinucleotide forms contacts with a specific subsite of IN recognizing the GT terminus of viral ds DNA. At the same time, specific ODNs containing CA or even only A at their 3′-termini probably are not in contact with the GT-recognizing subsite of the enzyme but interact with the second CA-recognizing subsite of IN. Data with nonspecific oligonucleotides also support this conclusion (Table 2). The affinity of short ODNs such as GGAA or $(GA)_3$ is comparable to that of specific CA-ODNs and $d(A)_n$, while the efficiency of interaction of AATT and $d(AC)_3$ correlates well with that of GT-ODNs and $d(T)_n$ of the same length. The same conclusion can be drawn from the change in the affinity of ds ODN_{21} (1.3 nM) after being modified; the removal of T from the 3′-end (ODN_{20}) or its replacement with rU (rU-ODN_{21}) decreased the affinity ~30−40-fold (Table 3B). The affinity of ds ODN_{20} (60 nM) or rU-ODN_{21} (50 nM) is similar to that of nonspecific ds $d(A)_{21}$·$d(T)_{21}$ (40 nM) of the same length, while ds CA-ODN_{19} presents a K_i comparable to that of the specific ds GT-ODN_{21} substrate.

Thus, these general trends in the interactions of IN with long and short ss ODNs corresponding to the cleaved strand of ODN_{21} agree very well with those for specific ds ODNs. The affinities of IN for ss GT-ODN_{21} (10 nM) and ss CA-ODN_{19} (30 nM) were nearly the same and only 7.7- and 2.5-fold lower, respectively, than those of their duplexes (Table 3). There is no doubt that the 3′-terminal T forms specific contacts with IN. In contrast, the interactions of the penultimate G are not so strong. The high affinity for GT-ODN_{21} and CA-ODN_{19} was expected since these ODNs have sequences representing the specific substrates for processing and strand transfer, respectively. After removal of GT, a reorganization of the active site of IN probably occurs, which allows effective interaction with the nascent 3′-terminal CA dinucleotide of the cleaved strand. Our results showing a stronger activation of IN by ODNs containing the 3′-terminal CA as compared with GT-containing ODNs also support this idea. Moreover, in this case, our data point to two different modes of interaction between IN and these ODNs.

The interaction of IN with the invariant 3′-terminal CAGT sequence of viral DNA either before or after the removal of GT contributes significantly to its total affinity for DNA. This sequence is very important for IN oligomerization into catalytically active forms and for cooperative interactions of the enzyme with the 3′-terminus and the 5′-flank of the specific ds ODN_{21} substrate. Moreover, the removal of GT leads to the formation of a free 3′-OH group on the terminal dA, which can form additional contacts, such as hydrogen bonds, with the amino acid residues of IN localized near this group. Interestingly, a replacement of dA with rA d(GAGATCGTC)rA (Table 2) decreases the affinity of the specific CA-ODN_{10} (Table 3) by a factor of ~5. This might result from formation of a hydrogen bond between two OH groups of the rA nucleoside, thus preventing the 3′-hydroxyl from having effective contacts with the enzyme. A decrease in the affinity of CTAGCA**p** by a factor of 6.5 after phosphorylation of the OH group from the deoxyribose of the terminal dA also suggests an important role of this OH group in the interaction with IN.

Affinity of IN for Specific DNA Duplexes. The data reported in Figure 2C show similar log-dependencies of K_i values versus length for nonspecific or specific double-stranded ODNs, although the affinity of IN for specific duplexes was always higher. The maximal contribution to the interaction between IN and specific DNA can be estimated approximately as $K_d = 0.033$ M (Table 3). The apparent relative contribution of the first strand (corresponding to the cleaved strand of viral DNA) to the overall affinity may be estimated as ~7−8 orders of magnitude, while adding the second strand increased the affinity by only ~1 order of magnitude. The same difference (~1 order of magnitude) corresponding to the relative contributions of the two complementary strands was observed for nonspecific ODNs. This indicates that IN interacts with both strands of the DNA substrate but that formally the contribution of the second strand is significantly lower. However, the increase in affinity because of the second strand and calculated from the respective affinities of ss and ds ODNs may be an underestimated value since the annealing of the second strand might induce changes in enzyme affinity.

Considering our results with HIV-1 integrase and our previous data (*9*, *10*) with DNA-specific enzymes (replication, repair, topoisomerization, and restriction enzymes), it can be concluded that the relative contribution of the second strand never exceeds 1−2 orders of magnitude. Different factors are involved, including partial melting of the DNA duplex in the complex with enzyme leading to a decrease in complementary interaction between strands. In other cases (e.g., in the case of a repair enzyme (*34*)), the protein forms more contacts with the first than with the second strand, and the relative contributions of these strands to the affinity of the DNA duplex usually correlate with the number of such contacts.

The reduced contribution of the second strand to the affinity does not mean that the second strand is not important for recognition and cleavage of specific viral DNA since IN cannot catalyze 3′-processing or integration with ss DNA. In the case of DNA repair, topoisomerization, and restriction enzymes, the second strand of DNA is usually more important for adjusting the DNA structure to a conformation optimal for catalysis than for enzyme binding. However, some enzymes cannot fit ss DNA to form direct contacts with the catalytic amino acid residues of the active site. The same might be true for HIV-1 IN.

Thermodynamic Models of IN Interactions with Specific DNA. The estimation of the relative contributions of different structural elements of specific and nonspecific DNA to the total affinity for any enzyme including IN seems difficult because of the dynamic and cooperative character of DNA−enzyme interactions.

In this study, we used the SILC approach to quantitate the determinants of substrate specificity. Within the limitations of the SILC method, the analysis of the data concerning the nucleotides from both strands of specific DNA to the

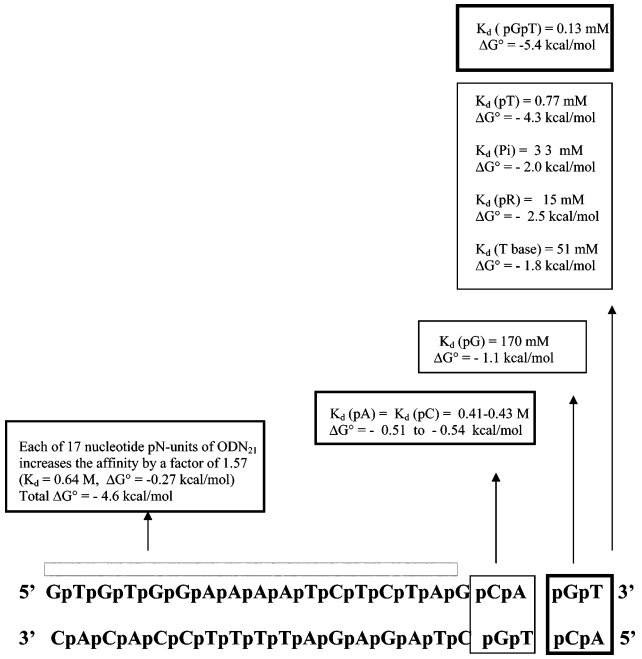

FIGURE 5: Thermodynamic model of the interaction of IN with GT-specific ds ODN_{21}. $\Delta G°$ values characterizing different contacts between the enzyme and the DNA strands are shown. All contacts of IN with ss GT-ODN_{21} provide an overall total $\Delta G°$ of approximately -11.1 kcal/mol ($K_d = 1 \times 10^{-8}$ M). The sum of all types of nonspecific and specific contacts of IN with the phosphate backbone and bases of the noncleaved strand together with the interactions between complementary strands, as compared to ss GT-ODN_{21}, increases the affinity approximately by a factor of 8 and decreases $\Delta G°$ by -1.25 kcal/mol. All types of nonspecific and specific interactions of IN and ds GT-ODN_{21} provide approximately $\Delta G° = -12.4$ kcal/mol.

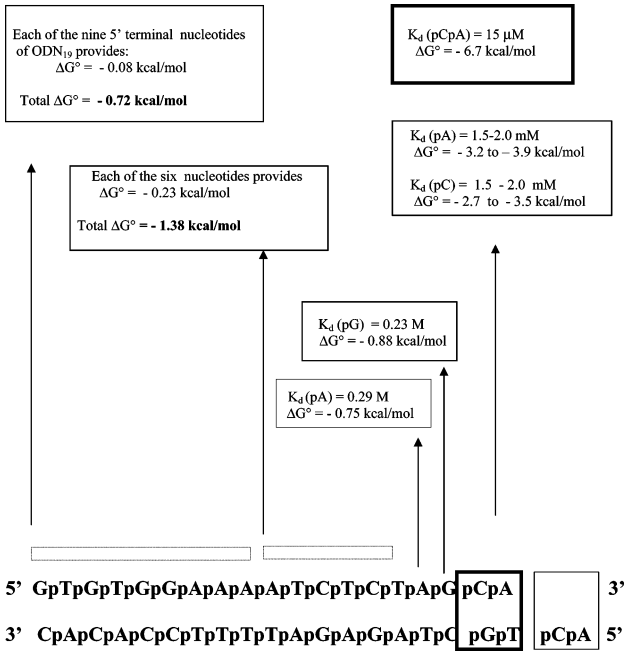

FIGURE 6: Thermodynamic model of the interaction of IN with the duplex CA-specific ODN_{19} and complementary 21-mer ODN. $\Delta G°$ values characterizing different contacts between the enzyme and the DNA strands are shown. All contacts of IN with the cleaved strand after removal of GT provide an overall total $\Delta G°$ of approximately -10.4 kcal/mol ($K_d = 3 \times 10^{-8}$ M). The sum of all types of nonspecific and specific contacts of IN with the phosphate backbone and bases of the noncleaved strand together with the interactions between complementary strands, as compared to ss ODN_{21}, increases the affinity approximately by a factor of 8 and decreases $\Delta G°$ by -1.25 kcal/mol. All types of nonspecific and specific interactions of IN and ds DNA provide approximately $\Delta G° = -11.6$ kcal/mol.

affinity of IN (Figure 2) showed that the relative contribution of the elements from the 3′-terminal nucleotides of the cleavable strand can differ by several orders of magnitude. In contrast, very small monotonic changes of the dependencies of log K_d on the number of nucleotides ($n > 4-8$) indicated weak interactions close to additive.

The relative contribution of GT to the affinity of GT-ODN_{21} for IN ($\Delta G° \approx -5.4$ kcal/mol) was approximately assessed from the K_d value (Table 3) for GT (Figure 5). The contributions of the phosphate, deoxyribose phosphate, and T nucleotide unit to the affinity of the GT-dinucleotide were obtained from K_d values of free Pi (0.033 M, $\Delta G° \approx -2.0$ kcal/mol), free d(pR) (0.015 M, $\Delta G° \approx -2.5$ kcal/mol), and free dTMP (0.77 mM, $\Delta G° \approx -4.3$ kcal/mol). Thus, the contributions of the base and the sugar of dTMP were estimated as $K_d \approx 0.051$ M ($\Delta G° \approx -1.8$ kcal/mol) and 0.450 M ($\Delta G° \approx -0.5$ kcal/mol), respectively. Since dCMP and dGMP have the same affinities as deoxyribose phosphate, these free nucleotides probably interact nonspecifically with a subsite of IN recognizing the T because of their pR moieties, similar to deoxyribose phosphate. Therefore, the approximate contribution of the G unit to the affinity of the GT dinucleotide for IN may be more correctly estimated from the ratio of K_d values for dTMP and the GT dinucleotide: K_d (G) $\approx$ 0.170 M ($\Delta G° \approx -1.1$ kcal/mol).

The relative contributions of the next A and C nucleotide units of the specific CAGT sequence are comparable and are characterized by the K_d values $\approx$0.41$-$0.43 M ($\Delta G° \approx -0.51$ to -0.54 kcal/mol). Each of the next 17 nucleotides of the 5′-flank increased the affinity by a factor of $\sim$1.57 ($K_d \approx 0.64$ M, $\Delta G° \approx -0.27$ kcal/mol). The total contribution of these 17 links was estimated as $\Delta G° \approx -4.6$ kcal/mol. All specific or nonspecific contacts of IN with the backbone and the bases of the noncleaved strand, together with the complementary interactions between the strands, increased the duplex affinity by a factor of 8 (corresponding to a decrease in $\Delta G°$ by -1.25 kcal/mol) as compared with ss ODN_{21}. Such a small increase in affinity may be due to several factors, including formation of weak contacts of IN only with the second strand of ds DNA or very weak contacts between the strands because of partial melting of ds DNA. Although the current data do not allow us to resolve this issue, the thermodynamic model of Figure 5 may approximately describe the interaction of IN with specific ds GT-ODNs.

For the processed substrate (CA-$ODN)_{19}$•$(N)_{21}$, the relative contribution of CA to the total affinity of CA-ODN for the enzyme ($\Delta G° \approx -6.7$ kcal/mol) was approximately evaluated from the K_d value for free CA (15 μM) (Figure 6). The independent estimation of the contributions of C and A units to the affinity of IN for the CA dinucleotide is rather difficult. Free dAMP and dCMP can interact with subsites of the enzyme different from those with which they are in contact when in the CA dinucleotide. In addition, the recognition of CA and AA dinucleotides by IN can occur in a cooperative fashion since the transition from dAMP to AA or CA led to a very large increase in affinity (by a factor of 63 and 333, respectively). Therefore, the minimal approximate contribu-

Table 4: Determination of K_M and V_{max} in the 3′-Processing Reaction Catalyzed by Integrase with Various ds ODNs Used as Substrates[a]

double-stranded ODNs	V_{max} (%)	K_M (nM)
[GTGTGGAAAATCTCTAGCAGT]·$(N)_{21}$	100	1.3
$[(T)_{17}$CAGT]·$(N)_{21}$	35	12
$[(A)_{16}$CAGT]·$(N)_{20}$	50	4.8
(-TGTGGAAAATCTCTA GCAGT)· ·(−ACACCTTTTAGAGAT **GAG**CA)	120	2.6
(-ACTGCTAGAGATTTTCCACA)· ·(-TGTGGAAAATCTCTA**CT** CGT)	not detectable	
$(T)_{19}$GT·$(N)_{21}$	not detectable	
$(T)_{21}$·$(A)_{21}$	not detectable	
$(T)_{17}$CCGT·$(N)_{21}$	not detectable	

[a] Errors of K_M and V_{max} values were within 20−30%. Values are the average results of two to three independent measurements.

tion of the A unit to the affinity of the dinucleotides can be estimated from K_d for free dAMP ($\Delta G° \approx -3.2$ kcal/mol). The extrapolation of the initial parts of biphasic curves for $d(A)_n$ and CA-ODN_{21} to $n = 1$ gives a K_i for the A unit ≈1.5−2 mM (−3.2 to −3.9 kcal/mol) for the first 3′-terminal nucleotide of these dinucleotides. Thus, a very approximate estimation of the contribution of the A and C units to the affinity of the CA dinucleotide is −3.2 to −3.9 and −2.7 to −3.5 kcal/mol, respectively.

The increase in affinity upon transition from the CA dinucleotide to longer ODNs occurs nearly additively, and the relative contributions of the G and A units (3rd and 4th from the 3′-end) to the affinity of CA-ODN was estimated as −0.88 and −0.75 kcal/mol, respectively (Figure 6). Each of the next six nucleotides of CA-ODN contributed approximately −0.23 kcal/mol to the total $\Delta G°$ of binding, their sum being −1.38 kcal/mol. Finally, the interaction of each of the nine 5′-terminal nucleotides of CA-ODN is characterized by $\Delta G° \approx -0.08$ kcal/mol, thus adding up to $\Delta G° \approx -0.72$ kcal/mol. All types of interactions of $(CA\text{-}ODN)_{19}$ with the complementary ODN_{21} increase the total $\Delta G°$ of binding by −1.2 kcal/mol. The thermodynamic model in Figure 6 might approximately describe the interaction of IN with ds $(CA\text{-}ODN)_{19}$·$(N)_{21}$.

Kinetic Factors: Reaction Rate and Specificity of Integrase. The sequence of the DNA substrate strongly influences the maximal rate of 3′-processing and integration reactions catalyzed by IN (*35*−*39*). We thus studied the substrate properties of some of the duplexes mentioned above in the 3′-processing reaction. As compared to the processing reaction obtained with the specific ds GT-ODN_{21}·$(N)_{21}$ substrate, no removal of dinucleotides from any single-stranded or double-stranded 21-mer ODNs was observed even at high IN concentration (5 μM) and incubation for 24 h (Table 4). However, partially specific ds ODNs such as $(T)_{17}$CAGT·$(N)_{21}$ or $(A)_{16}$CAGT·$(N)_{20}$, which contained the canonical 3′-terminal CAGT, were used as substrates by IN although at a reduced rate. In contrast, the V_{max} of the reaction was increased when the second strand was partially non-complementary, as in the case of GT-ODN_{21}. These results together with the data mentioned above point to an important role of duplex melting for productive interaction with IN.

Considering the high concentrations of nonspecific ODNs and IN at which there is no detectable reaction as compared with the rate of removal of 3′-terminal dinucleotides from specific ds ODNs at low concentrations of substrate and IN, it can be calculated that the reaction rate is increased by more than 5−6 orders of magnitude by the transition from nonspecific to specific ODNs. In contrast, the complex formation step accounts only for a ∼30-fold difference when changing from specific to nonspecific DNA. Thus, as in other enzymes, the catalytic step in integrase appears to be significantly more sensitive to the DNA structure than the IN·DNA complex formation step. The specificity of the enzyme action is mainly evidenced by the enzyme-dependent DNA adjustment to the optimal conformation and by the catalytic step of the reaction.

In conclusion, several steps of the reaction may provide enzyme specificity. First, in contrast to nonspecific DNA, specific DNA can direct the assembly of catalytically active oligomeric forms of IN and then activate the enzyme because of a change in its structure. Second, formation of complexes between preformed catalytically active oligomeric forms of IN and specific DNA occurs with approximately 1 order of magnitude higher efficiency as compared with nonspecific DNA. Upon transition from nonspecific to specific DNA, the k_{cat} of the processing reaction increases by >5−6 orders of magnitude. Finally, an additional increase in IN specificity may be provided by the particular stage of the integration reaction.

ACKNOWLEDGMENT

We are grateful to Ray Cooke (English Department, University Bordeaux 2) who edited the manuscript.

REFERENCES

1. Asante-Appiah, E., and Skalka, A. M. (1999) HIV-1 integrase: structural organization, conformational changes, and catalysis, *Adv. Virus Res. 52*, 351−369.
2. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration, *J. Virol. 63*, 5319−5327.
3. Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989) Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence, *Cell 58*, 47−54.
4. Craigie, R., Fujiwara, T., and Bushman, F. (1990) The integrase protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro, *Cell 62*, 829−837.
5. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro, *Cell 63*, 87−95.
6. Katzman, M., and Katz, R. A. (1999) Substrate recognition by retroviral integrases, *Adv. Virus Res. 52*, 371−395.
7. Katzman, M., and Sudol, M. (1998) Mapping viral DNA specificity to the central regions of integrase by using functional human immunodeficiency virus type 1/visna virus chimeric proteins, *J. Virol. 72*, 1744−1753.
8. Freemont, P. S., Lane, A. N., and Sanderson, M. R. (1991) Structural aspects of protein−DNA recognition, *Biochem. J. 278*, 1−23.
9. Nevinsky, G. A. (1995) Important role of weak interactions in long DNA and RNA molecules recognition by enzymes (a review), *Mol. Biol. (Moscow) 29*, 6−19.
10. Bugreev, D. V., and Nevinsky, G. A. (1999) Possibilities of the method of step-by-step complication of ligand structure in studies of protein−nucleic acid interactions: mechanisms of functioning of some replication, repair, topoisomerization, and restriction enzymes, *Biochemistry (Moscow) 64*, 237−249.
11. Lesser, D. R., Kurpiewski, M. R, and Jen-Jacobson, L. (1990) The energetic basis of specificity in the *Eco* RI endonuclease−DNA interaction, *Science 250*, 776−786.

12. Engler, L. E., Sapienza, P., Dorner, L. F., Kucera, R., Schildkraut, I., and Jen-Jacobson, L. (2001) The energetics of the interaction of *Bam*HI endonuclease with its recognition site GGATCC, *J. Mol. Biol. 307*, 619−636.
13. Engler, L. E., Welch, K. K., and Jen-Jacobson, L. (1997) Specific binding by *Eco*RV endonuclease to its DNA recognition site GATATC, *J. Mol. Biol. 269*, 82−101.
14. Jen-Jacobson, L. (1997) Protein−DNA recognition complexes: conservation of structure and binding energy in the transition state, *Biopolymers 44*, 153−180.
15. Holbrook, J. A., Tsodikov, O. V., Saecker, R. M., and Record, M. T., Jr. (2001) Specific and nonspecific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding, *J. Mol. Biol. 310*, 379−401.
16. Tsodikov, O. V., Holbrook, J. A., Shkel, I. A., and Record, M. T., Jr. (2001) Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer, *Biophys. J. 81*, 1960−1969.
17. van Tilborg, P. J., Czisch, M., Mulder, F. A., Folkers, G. E., Bonvin, A. M., Nair, M., Boelens, R., and Kaptein, R. (2000) Changes in dynamical behavior of the retinoid X receptor DNA-binding domain upon binding to a 14 base-pair DNA half site, *Biochemistry 39*, 8747−8757.
18. Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R., and Kaptein, R. (2002) Plasticity in protein−DNA recognition: *lac* repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain, *EMBO J. 21*, 2866−2876.
19. Caumont, A., Jamieson, G., Richard de Soultrait, V., Parissi, V., Fournier, M., Zakharova, O. D., Bayandin, R., Litvak, S., Tarrago-Litvak, L., and Nevinsky, G. A. (1999) High affinity interaction of HIV-1 integrase with specific and nonspecific single-stranded short oligonucleotides, *FEBS Lett. 455*, 154−158.
20. Ishchenko, A. A., Bulychev, N. V., Zharkov, D. O., Maksakova, G. A., Johnson, F., and Nevinsky, G. A. (1997) Isolation of 8-oxoguanine-DNA glycosylase from *Escherichia coli* and studies on its substrate specificity, *Mol. Biol.* (*Moscow*) *31*, 278−283.
21. Bodepudi, V., Iden, C. R., and Johnson, F. (1991) An improved method for the preparation of the phosphoramidites of modified 2′-deoxynucleotides: incorporation of oxo-2′-deoxy-7*H*-guanosine into synthetic oligomers, *Nucleosides Nucleotides 10*, 755−761.
22. Fasman, G. D., Ed. (1975) *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, Vol. 1, pp 589, CRC, Boca Raton, FL.
23. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.
24. Fersht, A. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman, New York.
25. Cornish-Bowden, A. (1976) *Principles of Enzyme Kinetics*, pp 160−206, Butterworth, London and Boston.
26. Tuzikov, F. V., Zinoviev, V. V., Vavilin, V. I., and Malygin, E. G. (1996) Application of the small-angle X-ray scattering technique for the study of the two-step equilibrium enzyme−substrate interactions, *Biopolymers 38*, 131−139.
27. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
28. Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8233−8238.
29. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) Predicting DNA duplex stability from the base sequence, *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
30. Asante-Appiah, E., and Skalka, A. M. (1997) A metal-induced conformational change and activation in HIV-1 integrase, *J. Biol. Chem. 272*, 16196−16205.
31. Yi, J., Asante-Appiah, E., and Skalka, A. M. (1999) Divalent cations stimulate preferential recognition of a viral DNA end by HIV-1 integrase, *Biochemistry 38*, 8458−8468.
32. Deprez, E., Tauc, P., Leh, H., Mouscadet, J. F., Auclair, C., and Brochon, J. C. (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy, *Biochemistry 39*, 9275−9284.
33. Tuzikov, F. V., Tuzikova, N. A., Panin, L. E., Nikitin, Y. P., and Krylova, I. N. (1998) Determination of the fraction composition of blood lipoproteins by the small-angle X-ray scatterng technique, *Membr. Cell. Biol. 12*, 521−536.
34. Ishchenko, A. A., Vasilenko, N. L., Sinitsina, O. I., Yamkovoy, V. I., Fedorova, O. S., Douglas, K. T., Nevinsky, G. A. (2002) Thermodynamic, kinetic, and structural basis for recognition and repair of 8-oxoguanine in DNA by Fpg protein from *Escherichia coli*, *Biochemistry 41*, 7540−7548.
35. Leavitt, A. D., Rose, R. B., and Varmus, H. E. (1992) Both substrate and target oligonucleotide sequences affect in vitro integration mediated by HIV-1 integrase protein produced in *Saccharomyces cerevisiae*, *J. Virol. 66*, 2359−2368.
36. Katzman, M., and Sudol, M. (1996) Influence of subterminal viral DNA nucleotides on differential susceptibility to cleavage by human immunodeficiency virus type 1 and visna virus integrases, *J. Virol. 70*, 9069−9073.
37. Balakrishnan, M., and Jonsson, C. B. (1997) Functional identification of nucleotides conferring substrate specificity to retroviral integrase reactions. *J. Virol. 71*, 1025−1035.
38. Brown, H. E., Chen, H., and Engelman, A. (1999) Structure-based mutagenesis of the HIV-1 DNA attachment site: effects on integration and cDNA synthesis, *J. Virol. 73*, 9011−9020.
39. Zhou, H., Rainey, G. J., Wong, S.-K, and Coffin, J. M. (2001) Substrate sequence selection by retroviral integrase, *J. Virol. 75*, 1359−1370.

BI0300480