Chapter 5

Preface
5.1 Introduction

This section documents studies in molecular recognition using a model system involving the structural analysis of functional mimicry between peptides and nucleic acids. Insight into the principles of molecular recognition is as crucial to the development of better structure prediction and design methods as an understanding of macromolecular assembly and stability, an issue examined in the previous chapters. Intensive research activity has focused on this area since insight into these principles opens the possibility of rationally designing complex systems capable of selective recognition and catalysis as well as developing new drug therapies. However, intermolecular interaction phenomena continue to remain poorly understood and design from first principles remains an elusive goal. One of the possible reasons for this could be that the full repertoire of non-covalent interactions that govern recognition events is yet to be fully defined. For instance, the importance of interactions such as those between ions and the π-ring have not been fully appreciated in biological systems. Also, the importance of the ways in which individual non-covalent forces either compete and negate or reinforce each other in complex systems is often underrated. For instance, the contribution of a hydrogen bond to the stability of binding may often be overestimated by overlooking the destabilizing effects of torsional strain.

Given the success of large-scale genome sequencing projects, an understanding of the networks of molecular interactions that govern and coordinate functions, both at the cellular level and at the level of the organism, has become imperative. Therefore, numerous molecular interaction databases, which organize and allow easy access to available binding data have taken shape (Gilson, 2002). Each of these databases focuses on the different experimental
aspects of binding studies. Thus Relibase focuses on the 3D structures of proteins with small organic molecules such as co-factors and drug candidates and comprises of data derived from the PDB (Hendlich, 1998; Bergner et al., 2002). Relibase allows the retrieval and analysis of protein-ligand interaction patterns with 3D constraints, such as favorable combinations of interacting functional groups or preferred interaction geometry. The Biomolecular Interaction Network Database (BIND) focuses primarily on identifying protein-protein interactions, but also annotates interactions involving DNA, RNA and small molecules (Bader et al., 2001; Salama et al., 2002). In addition to extracting information from 3D structures (Salama et al., 2002), BIND also draws data from yeast two-hybrid studies, mass spectrometry and other sources. The Thermodynamic Database for Protein-Nucleic Acid Interactions (ProNIT) is comprised of a collection of binding affinities and associated data on protein-nucleic acid binding (Sarai et al., 2002). The Binding Database (Binding DB) compiles information about the affinities of biomolecular association including those involving synthetic compounds (Chen et al., 2002). As with DNA and protein sequence databases, which have assisted enormously in inferring possible structural and functional roles, it is hoped that the expansion and development of these databases will eventually facilitate the creation of suitable knowledge-based potentials, which allow the simulation of molecular interactions.

5.1.1 Characteristics of Binding Events

Extensive studies of binding surfaces on proteins and other biopolymers have highlighted the importance of shape and charge complementarity of the interacting surfaces. However, in several instances, molecules are known to drastically reorganize their binding sites to optimize the binding interaction.
Interactions between molecules may also be mediated by water, as is the case with protein-DNA and protein-carbohydrate binding, where water serves to generate a structurally complementary surface. The specificity or fidelity of binding exhibits enormous variations. While some molecules, like albumin, can bind to a wide variety of ligands, others such as the t-RNA synthetases can distinguish subtle differences between single amino acids.

5.1.1.1 Atomic Structure of Interaction Sites

Numerous surveys of biomolecular interaction surfaces derived from the PDB have been carried out, and these reveal broad structural similarities. In dimeric protein-protein complexes, the size of the interface area per subunit has been found to vary from 368 to 4890 Å² (Janin et al., 1988; Jones & Thornton, 1995), which represents roughly 12% of the accessible surface area (Argos, 1988). The interface area for higher oligomers increases, with monomeric subunits of tetramers averaging 20.9% of the total surface area (Janin et al., 1988). The interface area of proteins in general have been found to be more hydrophobic than the rest of the protein exterior but more polar than the protein interior (Janin et al., 1988; Korn & Burnett, 1991; Tsai et al., 1997). This is justified by the need to promote association without destabilizing the isolated monomer. Along similar lines, Young et al. (1994), in a survey of protein-protein and protein-peptide complexes have found that the most hydrophobic solvent-accessible cluster of residues formed part of the intermolecular interface in a significant proportion of the complexes. Interfaces also exhibit a marked preference for certain amino acids, with a disproportionately high occurrence of arginine, asparagine and tyrosine residues (Janin et al., 1988; Argos, 1988; Janin & Chothia, 1990; Jones & Thornton, 1995). The average number of interface residues in a
survey of protease-inhibitor and antigen-antibody complexes has been found to be 34±7 (Janin & Chothia, 1990). Several studies have reported that the amount of solvent-accessible surface buried in oligomeric proteins is related to the molecular weight of the monomer by a simple power law (Miller et al., 1987a). A similar relationship has been found to hold true for interactions involving monomeric proteins (Bryant et al., 1989; Miller et al., 1987b). These empirical relationships allow an approximate determination of the interface area involved in protein interactions. Lawrence and Colman (1993) have defined the shape correlation statistic as a measure of shape complementarity of protein-protein interfaces. Their study together with some others (Jones & Thornton, 1995, 1996) has found that while dimeric protein complexes have most complementary interfaces, protease-inhibitor complexes are less so and antibody-protein antigen complexes are least complementary.

Extensive work has been carried out towards defining the molecular basis of protein-DNA and protein-RNA recognition. While the broad diversity of RNA structures makes the specific recognition of unique shapes and charge distributions of different RNA sequences possible, the relative absence of structural diversity in DNA necessitates a direct readout of individual nucleotides for sequence-specific recognition. Analyses of DNA-binding sites have revealed numerous shared features (Jones et al., 1999). These sites are largely populated by polar residues, in sharp contrast to protein-protein interfaces, and are capable of making numerous direct and water-mediated hydrogen bonds. The DNA-binding proteins bury roughly $618 \text{Å}^2$-$2833 \text{Å}^2$ of their accessible surface during binding. Their interfaces may be comprised of 2 to 16 discontinuous segments, with the dimeric protein interfaces being more
segmented than the monomeric proteins. Enzymes tend to have larger and more highly segmented protein-DNA interfaces than transcription factors. The structural parameters for bound DNA (including twist, slide and roll) have been found to be intermediate of those observed in unbound B-DNA and A-DNA. Jones et al. (1999) have identified three basic modes of protein binding to DNA: single-headed, double-headed and enveloping. Proteins that contact DNA using a single or double cluster of residues have been classified as single- or double-headed. Those proteins where DNA base- or backbone-contacting residues reside in a cleft or deep groove have been defined as enveloping. While DNA-binding enzymes interact by enveloping DNA, those that function as transcription factors contact DNA with either one or two binding clusters. Most DNA-binding proteins make contacts in the major groove of DNA. The eukaryotic transcription factors bind by the insertion of an α-helix into the major groove of DNA. Many transcription factors, including the leucine zipper and the homeodomain proteins share sequence features within a 12-residue segment along this insertion helix (Suzuki, 1993; Suzuki, 1994). While Arg/Lys residues at certain positions in the C-ter half of this segment contact phosphate groups, other residues at positions in the N-ter half interact with bases.

In a more recent study, Luscombe et al. (2001) have analyzed a set of 129 protein-DNA complexes in an effort to formulate a more generic code for sequence-specific DNA recognition. They report that roughly two-thirds of all protein-DNA interactions involve van der Waals contacts, and one-third are contributed by direct and water-mediated hydrogen bonds. Over two-thirds of the contacts are made with the sugar-phosphate backbone of DNA, which nonspecifically stabilize the protein-DNA complex. Base contacts are established primarily in the major groove. Approximately two-thirds of the hydrogen bonds
with bases are of the bidentate (where two or more hydrogen bonds are made with a base or base pair by a single amino acid residue), or complex (where a protein residue binds more than one base step simultaneously) type. The hydrogen bond distribution clearly demonstrates that certain amino acid-base pairs are favored. Thus arginine, lysine, serine and histidine preferentially interact with guanine while asparagine and glutamine prefer adenine. This observation is also supported by previous studies (Suzuki, 1994; Mandel-Gutfreund et al., 1995). Complex interactions, where multiple base steps are recognized by the same residue, make discriminatory sequence-specific recognition possible. Such interactions are expected to play a crucial and possibly generic role in binding specificity.

Sequence-specific RNA recognition occurs by binding to single-stranded regions and hairpin loops or those sites in double-helical regions where internal loops or bulges induce structural distortions. The major groove of double-helical RNA is too narrow to allow the insertion of an α-helix or β-strand. Most RNA-binding proteins possess modules for RNA recognition. Of these, the double-stranded RNA-binding domain (dsRBD) binds double-stranded RNA with little or no sequence specificity, though multiple domains may specifically recognize certain RNA structures (Ferrandon et al., 1994; Bycroft et al., 1995; Kharrat et al., 1995). KH (K-homology) domains are non-specific single-stranded RNA-binding proteins (Musco et al., 1996) whereas RNP (ribonucleoprotein) domains specifically recognize both single-stranded and highly structured RNAs (Nagai et al., 1990).

Carbohydrates play diverse roles in mediating biological recognition events. A majority of the protein-carbohydrate structures available are of legume lectins such as concanavalin A, pea, lentil and other lectins, and the periplasmic
proteins such as the L-arabinose, glucose, galactose and maltose binding proteins. The crystal structures of several available legume and cereal lectins complexed to oligosaccharide ligands (Rini et al., 1993; Delbaere et al., 1993; Wright & Jaeger, 1993; Bourne et al., 1994) reveal several common features (Toone, 1994; Drickamer, 1997). One recurring theme is the presence of several water molecules that mediate protein-carbohydrate hydrogen bonding contacts and serve as molecular mortar at the protein-carbohydrate interface. Carbohydrate ligands can be broadly classified into two groups based on the way these are recognized by lectins. One group, which includes glucose, mannose and N-acetylglucosamine, is distinguished by an equatorial arrangement of 3- and 4-hydroxyl groups. The second group, which includes galactose and N-acetylgalactosamine, is characterized by an axial arrangement of the 4-hydroxyl group. This difference serves to impart specificity to lectin binding. In general, van der Waals interactions and hydrogen bonds stabilize and orient the sugar in the binding site. Once an orientation is established, selectivity is achieved by interactions with those hydroxyl groups that distinguish favored from unfavored ligands, and also by steric exclusion. In structures of the S-type animal lectins (Lobsanov et al., 1993; Liao et al., 1994), the saccharide hydroxyl groups participate in extensive hydrogen bonding networks with acidic side chains. In addition, the complexes are stabilized by van der Waals contacts with hydrophobic residues, and numerous water-mediated hydrogen bonds are also visible. The molecular mechanisms of antibody-carbohydrate antigen binding have been examined with detailed comparisons to lectin-sugar binding (Cygler et al., 1991; Rose et al., 1993; Bundle et al., 1994). The carbohydrate-binding grooves are usually lined with aromatic residues, and the antibody relies heavily on hydrophobic van der Waals contacts to bind sugars in
sharp contrast to lectin-sugar binding. In the crystal structure of β-amylase bound to cyclodextrin (Mikami et al., 1993), the ligand makes remarkably few hydrogen bonding interactions with the protein, and the complex is stabilized largely by van der Waals contacts with nonpolar residues. In the crystal structures of D-xylose isomerase bound to glucose and 3-O-methyl glucose (Lavie et al., 1994), the glucose molecule interacts with two Mg\(^{2+}\) ions in the enzyme active site, and is also hydrogen bonded to histidine and lysine residues. A common feature between protein-carbohydrate and protein-DNA interactions is the presence of bridging water molecules that mediate hydrogen bonding. These water molecules not only generate a structurally complementary surface, but in many protein-DNA complexes, by mediating interactions between acidic amino acid residues and the sugar-phosphate backbone, screen the forces of electrostatic repulsion.

Crystallographic analyses of steroid-binding proteins, enzymes and antibodies have helped elucidate some of the principles of steroid recognition. Antibodies have been found to bind steroids in nonpolar grooves or pockets in the complementarity determining regions (Arevalo et al., 1993, 1994; Jeffrey et al., 1993, 1995). Comparison of both complexed and uncomplexed receptors reveal highly preorganized binding sites. The antibody-steroid antigen interaction is characterized by high host-guest complementarity and the burial of large nonpolar regions of the substrate on the order of 220-350 Å\(^2\). Aromatic amino acid side chains account for 50-60% of the nonpolar contacts with the ligand. Hydrophobic desolvation and dispersion are the primary forces driving association. In the crystal structures of cholesterol oxidase (Vrielink et al., 1991; Li et al., 1992) and the steroid dehydrogenases (Hoog et al., 1994; Ghosh et al., 1994a,
1994b, 1995) complexed with cognate steroid ligands and steroidal inhibitors, the ligands are either bound in hydrophobic cavities or in nonpolar regions at the interface of two or more monomeric subunits. Unlike anti-steroid antibodies, aromatic amino acid side chains appear less important for complex formation, and there appears to be some flexibility in terms of ligand binding and association.

5.1.1.2 Thermodynamics and Kinetics of Recognition Events

While the elucidation of the structural details of inter-molecular association is essential to understanding the principles of recognition, it represents only half the picture. A quantitative estimation of the forces of association is just as essential to obtain a unified picture of structure and function. The thermodynamics of protein-protein and protein-peptide association have been extensively studied by various groups (Baugh & Townbridge, 1972; Tello et al., 1994; Martinez et al., 1995; Jelesarov & Bosshard, 1996; Petrella et al., 1996). While some of these studies support enthalpy, others contradictorily support entropy as the driving force for association. In studies carried out 25°C, enthalpy has been found to drive complex association in 74% of the cases, while entropy favors complex formation in 55%. The average ΔG for protein-protein association has been found to be -10.4 kcal/mol whereas the average ΔG for protein-peptide interactions is -8.5 kcal/mol. The average ΔH and ΔS for protein-protein interactions are -8.6 kcal/mol and 6.12 cal/(mol K) respectively, and for protein-peptide interactions are -8.9 kcal/mol and -1.13 cal/(mol K) respectively. The mean ΔCp for protein-protein and protein-peptide interactions are -333 and -447 cal/(mol K) respectively. Large negative values of ΔCp are characteristic of the burial of hydrophobic groups during a process (Gill
& Wadsö, 1976; Spolar et al., 1992) whereas smaller positive values characterize the burial of polar groups in a nonpolar environment (Murphy & Gill, 1991). The mean values of $\Delta C_p$ suggest that the hydrophobic effect is the primary force driving protein-protein association.

The hydrophobic effect, along with the release of cations, is considered the major force driving protein-DNA complex formation (Ha et al., 1989). Calorimetric studies on diverse protein-DNA complexes, including the binding of the Cro repressor (Takeda et al., 1992), trp repressor (Ladbury et al., 1994), met repressor (Cooper et al., 1994; Hyre & Spiceer, 1995), and GRDBD (Lundbäck et al., 1994; Lundbäck & Härd, 1996), with their cognate DNA targets, indicate the opposing effects of temperature on $\Delta H$ and $-T\Delta S$. Therefore, $\Delta G$ is relatively insensitive to temperature changes. However, this also implies that at lower temperatures the entropic contribution to $\Delta G$ dominates while at higher temperatures the enthalpic contribution dominates. Thus at lower temperatures protein-DNA association is entropy driven whereas at higher temperatures, enthalpy drives this process. Complex formation is accompanied by a large negative change in heat capacity. This has been attributed to the release of water molecules during complexation as well as increased rigidity of the protein interface (Spolar & Record, 1994; Von Hippel, 1994), and of the water molecules residing at the interface.

Calorimetrically determined thermodynamic constants for protein-carbohydrate association, in almost all cases, indicate that enthalpy is the driving force for binding (Williams et al., 1992; Schwarz et al., 1993; Bundle et al., 1994; Toone, 1994). The thermodynamic data suggest a strong linear enthalpy-entropy compensation with a slope slightly greater than unity (Toone, 1994). This has
been interpreted in terms of changes in the rotational degrees of freedom as glycosidic torsion angles are frozen, and in terms of solvent reorganization accompanying binding. The values of $\Delta C_p$ for lectin-carbohydrate binding are small and negative whereas those for antibody-carbohydrate association are large and temperature dependent (Sigurskjold & Bundle, 1992). This disparity has been interpreted in terms of the structural differences between lectin- and antibody-binding sites. The antibody-binding site contains several aromatic amino acids, and a large $\Delta C_p$ reflects the importance of the expulsion of solvent during complex formation.

5.1.2 Approaches to the Problem

Studies on molecular recognition have been carried out by different methods and on varied systems, and these have elucidated various aspects of the mechanism of recognition. Some of these are discussed below.

5.1.2.1 Mutational Analyses

Examining the effects of mutating residues between interacting protein surfaces has been a standard approach for probing protein-protein interactions. While many of these studies are aimed at the disruption of the interface (for instance the introduction of isolated charges or ion pairs) to examine protein-protein association, others involve a systematic mutation at the interface in order to examine the contribution of specific residues. Thus tetrameric fructose-1,6-bisphosphate aldolase has been converted into a dimeric form by the substitution of an aspartate at the interface with other residues (Beernink & Tolan, 1994). A series of neutral and charged residue substitutions at ProB28 and LysB29 in insulin are known to disrupt oligomer formation (Shoelson et al., 1992; Brems et al., 1992). The association of human growth hormone with its receptor has been
the subject of numerous studies by alanine scanning mutagenesis (Cunningham & Wells, 1989, 1993; Bass et al., 1991; Clackson & Wells, 1995). An unusually small number of the contact residues on the hormone or the receptor have been found to be energetically significant. The energetically important residues in the receptor form a hot spot together with the energetically important residues on the hormone, and this hot spot is largely hydrophobic. In another thermodynamic study analyzing the binding of factor VIIa to the extracellular domain of human tissue factor through alanine scanning mutagenesis, a small number of interfacial residues have been found to largely contribute to the binding affinity (Kelley et al., 1995). The importance of a small number of interfacial residues for the binding reaction has been found to be a general trend in many studies involving different proteins. Another general conclusion drawn from numerous studies has been that the loss of binding energy accompanying an interfacial mutation is primarily due to an increase in the dissociation rate, rather than any change in association rates. One specific instance where extensive mutational analyses coupled with crystallographic studies has helped elucidate a clear one-to-one amino acid-base recognition code is in the binding of zinc fingers to DNA (Choo & Klug, 1997). Some of these rules have subsequently been used to endow zinc fingers with desired specificities (Desjarlais & Berg, 1993).

5.1.2.2 Model Systems

Biomolecular interactions involve complex patterns of size, shape and chemical complementarity between interacting surfaces. Attempts have been made to replicate such behaviour in simpler model systems that are synthetically accessible. Though these models are still very primitive compared to biological
systems, some level of sophistication has been achieved. 'Host molecules' that feature a cleft have been extensively used for designing complementary shapes since functional groups attached to the interior of the cleft converge on 'guests' held within. Large macrocyclic compounds such as cyclodextrins, cyclophanes and crown ethers are particularly well suited for this purpose. The cyclodextrins have internal dimensions which are large enough to incorporate aryl units, and bind to a variety of compounds that include the porphyrins and the steroids (Conn & Rebek, 1997; Wallimann et al., 1997). The cyclophanes have similar shapes, and in addition to hydrophobic binding provide directionality in their interactions with the substrate by aromatic ring stacking (Rebek, 1987). The crown ethers act primarily on ionic substrates, and bind to various cations with modest levels of specificity (Cram, 1983; Rebek, 1987).

Significant advances have been made in the design of receptors for linear peptides. Macrocyclic cage-like receptors consisting of 1,3,5-triarylbenzene spacers connected by amino acid linkers have been designed to bind short peptides and N-protected amino acids (Pieters & Diederich, 1996; Dowden et al., 1997). The peptide linkers provide hydrogen bonding sites for the substrates with potential enantioselective recognition due to their inherent chirality. Another form of peptide recognition relies on the formation of β-sheets (Nowick et al., 1996, 1997). This can be accomplished by the incorporation of a β-turn mimic into a peptide chain or an extended peptide mimic which inter- or intramolecularly induces β-sheet formation in another peptide.

Various strategies have been adopted for designing receptors that bind to carbohydrates. One of these involves receptors bearing phosphonate groups that bind to various glycosides with modest enantioselectivity (Das & Hamilton,
The anionic phosphonate groups form hydrogen bond acceptors for glycosidic hydroxyl groups. In addition, the cyclophanes and the cyclodextrins have also been explored as potential receptors for glycopyranosides (Staley & Smith, 1996; Junquera et al., 1996). Extensive calorimetric studies on the interactions of α-cyclodextrin with a series of p-nitrophenyl glycopyranosides have been carried out, and these suggest that hydration effects and van der Waals interactions are the primary forces driving association (Junquera et al., 1996).

Significant advances have been made in the design of synthetic ligands that are capable of binding to nucleic acids with modest levels of sequence specificity. Oligonucleotides, which bind to DNA by Hoogsteen base pairing and triple helix formation have been used to alter gene expression (Maher et al., 1992; Duvalentin et al., 1992). The design of numerous minor groove binding polyamides has formed an alternative approach towards sequence specific recognition. Simple empirical rules have been developed for specific minor groove binding using polyamides containing the amino acids imidazole and pyrrole, and their derivatives (Trauger et al., 1996; White et al., 1998). These constitute a promising class of molecules for high affinity sequence specific DNA recognition.

5.1.2.3 Computational Methods

A number of computational methods have been developed to solve the problem of defining binding sites as also the characteristics of binding. The MCSS (multiple copy simultaneous search) method allows small probe molecules to interact with a target protein and subsequently determines potential energy minima to define favorable interaction sites (Miranker & Karplus, 1991).
A number of algorithms that predict the docking of two proteins utilize an initial rigid-body global search for surface complementarity followed by subsequent screening and refinement. The DOCK algorithm (Shoichet & Kuntz, 1996) follows the rigid-body docking approach, and fills the active site of a protein with a cluster of overlapping spheres. The algorithm matches the sphere centers of this cluster with those of potential ligands followed by a ranking of complexes based on contact scores or potential energy.

Flexible protein docking strategies involve Monte Carlo and simulated annealing methods incorporating prior knowledge of the ligand binding site thereby narrowing the search space (Hart & Read, 1992; Stoddard & Koshland, 1993).

5.1.2.4 Analysis of Mimicry

Examining the number and nature of interactions in systems where functional equivalence exists in the absence of chemical similarity can serve as a useful tool for elucidating structure-function relationships. Examples of molecular mimicry abound in biological systems. Overall structural similarities between the elongation factors EF-Tu and EF-G within the protein biosynthetic apparatus have been well documented (Kjeldgaard et al., 1993; Berchtold et al., 1993; Aevarsson et al., 1994; Czworkowski et al., 1994). Both proteins are structurally related with conformationally similar guanine nucleotide-binding N-ter domains, while three domains at the C-ter of EF-G resemble tRNA, and relative to EF-G's nucleotide binding site are positioned in the same way as tRNA in the ternary complex of EF-Tu, GTP and aminoacylated tRNA. Both EF-Tu and EF-G compete for the same binding site on the ribosome, and the resemblance of EF-G to the ternary complex of EF-Tu, GTP and aminoacylated
tRNA is suggestive of mechanistic similarities between EF-G catalyzed translocation and EF-Tu catalyzed tRNA binding to the ribosome. Another example is the generation of anti-idiotypic antibodies during an immune response bearing the internal image of the original epitope. Among numerous known examples of this phenomenon, anti-idiotypic antibodies produced in response to insulin immunization have been shown to bind the insulin receptor and induce glycolysis (Gaulton & Greene, 1986).

The system employed for studies documented herein involved the use of peptide inhibitors of RNase A that competitively inhibited its activity. RNase A preferentially hydrolyzes single-stranded RNA on the 3'-side of a pyrimidine base. Extensive structural studies of RNase A have been carried out in complex with substrate analogues, transition state analogues, products and inhibitors (McPherson et al., 1986a, 1986b; Birdsall & McPherson, 1992; Fontecilla-Camps et al., 1994; Zegers et al., 1994; Aguilar et al., 1991, 1992). Structural and biochemical data reveal the existence of several enzymic subsites which interact with bases and phosphate groups of the bound substrate. The organization of the substrate-binding site is represented as a schematic in Figure 5.1. Of the three characterized enzymic subsites for bases, the B1 subsite preferentially binds pyrimidines while the B2 and B3 subsites prefer adenine and purine respectively. The enzyme catalyzes the cleavage of the phosphodiester bond of the phosphoryl group bound at the P1 site (Figure 5.1).

A comparative structure-function analysis of RNase A-peptide complex formation with available data on RNase A-nucleic acid interactions should help elucidate the features associated with peptide-nucleotide mimicry. This should be potentially useful in the design of ligands for RNase A that do not possess the
Figure 5.1 Schematic diagram illustrating the RNase A-substrate interaction. B and p indicate binding sites for base and phosphate moieties respectively. B1 is specific for pyrimidines while B2 preferentially binds purines.
nucleic acid scaffold. Efforts in this regard are discussed in the following chapters.