Molecular recognition and the development of self-replicating systems

J. Rebek Jr

Department of Chemistry, The Massachusetts Institute of Technology, Cambridge (Massachusetts 02139, USA)

Abstract. Weak intermolecular forces lie at the heart of biochemical recognition phenomenon and the last decade has seen much activity in the evaluation of these forces. A number of model systems have been developed including macrocyclic structures and molecular clefts. With these structures it has been possible to measure forces at the sub-kilocalorie level involving hydrogen bonding, aromatic stacking and van der Waals interactions. This manuscript deals with molecular clefts as synthetic receptors for nucleic acid components and their ultimate use in developing chemical reactions between components within a complex. This has led to an entirely synthetic, self-replicating system that shows the features of self-complementarity and autocatalysis. A general discussion of self-replicating systems and their implications for prebiotic chemistry is developed.

Key words. Adenine; aryl stacking; autocatalysis; base-pairing; catalysis; hydrogen bonding; intermolecular forces; nucleic acids; recognition; replication; self-complementarity.

The last decade saw a sharp increase in activity in molecular recognition; weak intermolecular forces became one of the paradigms of bioorganic chemistry. During this time advances in enzymology permitted measurement of the contributions of hydrogen bonding and hydrophobic effects in enzyme substrate interactions[11]. The use of model systems for these purposes has led to methods that provide a dissection of intermolecular forces at the sub-kilocalorie level. In this review, we discuss the application of these systems to problems involved with self-replication.

A classic form of molecular recognition is the base pairing that occurs in double-stranded nucleic acids. Genetic information is transmitted in the horizontal direction (fig. 1) through hydrogen bonds, whereas additional stability in the form of aromatic stacking interactions is conferred in a vertical direction to the double-stranded forms of nucleic acids[18]. Early model studies showed considerable stacking type aggregations occur in water[4], since hydrogen bonding between single base pairs cannot be detected in this medium. Accordingly, initial studies were performed in $CHCl_3$ by Rich[14], using IR techniques; more recently, NMR has been used for this purpose[23].

It is now possible to construct molecules in which simultaneous stacking and base pairing can be observed. In such models these forces converge from perpendicular directions and imitate the geometries of the naturally occurring systems. Molecular shape becomes the factor, and we were fortunate in having access to molecules with which advantageous shapes could be sculpted in rapid and modular fashion.

Specifically, the Kemp[12] triacid features a U-shaped relationship between any two carboxyl groups (fig. 1). As a result, molecules in which it is used as a module 'fold back' upon themselves. Condensation of this material

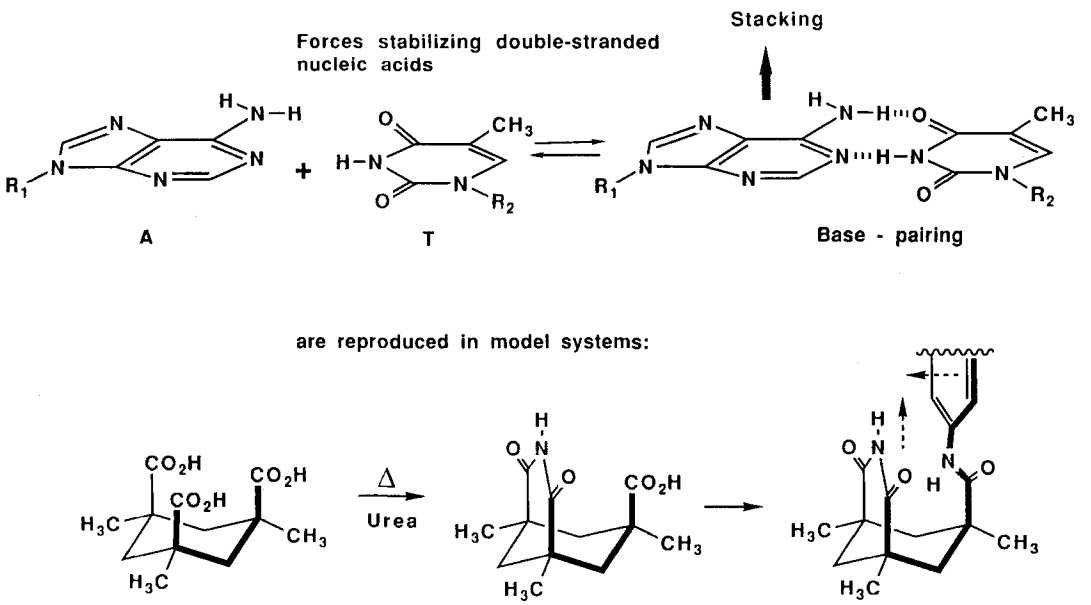

Figure 1

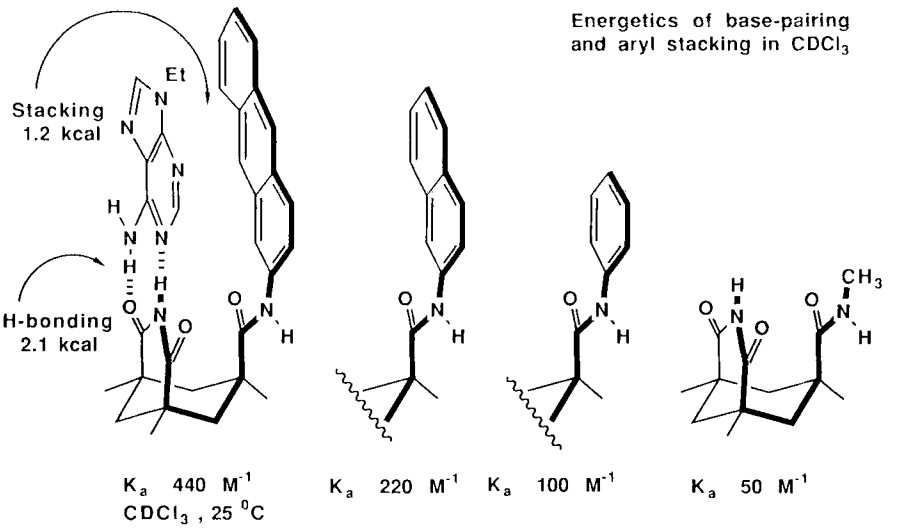

Figure 2

with urea gives an imide that provides a hydrogen bonding edge similar to that of thymine. Virtually any aromatic surface can be appended to the third carboxyl group by means of an ester or amide linkage[17]. It is therefore possible to make molecules that show high affinity for adenine derivatives through simultaneous hydrogen bonding and stacking interactions. For example, the series featuring methyl, phenyl, naphthyl and anthracyl surfaces were prepared, and association constants (K_a) with soluble adenine derivatives were measured[22] in the relatively non-competing solvent, $CDCl_3$. Within this series, the hydrogen bonding contribution remains constant, and extraction of the effects of aromatic stacking is possible. About one kilocalorie in binding energy can be obtained from the dipole-induced dipole interactions that result when an adenine is brought near an anthracyl surface. These forces are quite distinct from the classical or nonclassical[6] hydrophobic effects in which solvent liberation plays a large role.

Such model structures are also useful in exploring the details of the base-pairing or hydrogen-bonding events. Using Nuclear Overhauser Effect (NOE) techniques, it is possible to show that typical A-T base pairing in this solvent occurs with roughly equal amounts of Watson-Crick and Hoogsteen configurations[2]. In the aromatic derivatives, however, it is possible to control the relative contributions of both modes. For example, with the longer, extended anthracyl surface (e.g., fig. 2), Watson-Crick base pairing is favored, since more efficient overlap of the aromatic π systems occurs. Alternatively, by introducing remote steric effects on the periphery of the structures such as shown for the t-butyl naphthyl derivative (fig. 3), it is possible to force the base pairing to be predominantly Hoogsteen. Only in this manner is it possible to bring the two aromatic surfaces into contact without causing unpleasant interactions among the side chains. Heteronuclear NOE techniques could also be applied to this system, and these revealed that bifurcated hydrogen bonds exist which confer additional stability to the complexes. For example, the amide (fig. 4) shows contact of the NH of adenine with all carbonyls in the system[16], whereas the corresponding ester shows only hydrogen bonding at the imide function. The enhancement due to

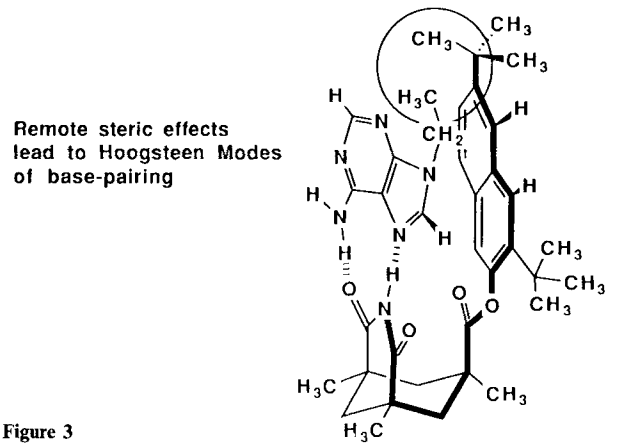

Figure 3

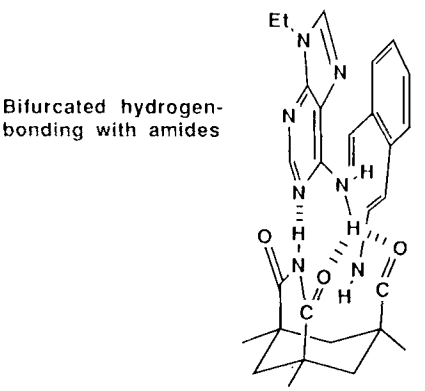

Figure 4

Chelation and transport of adenine derivatives

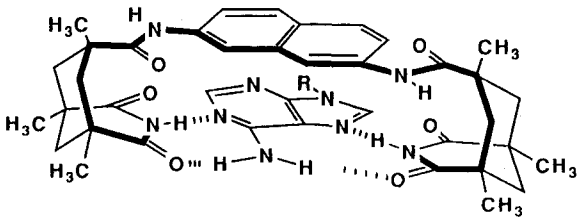

Figure 5

the bifurcation appears to be a factor of about 0.4 kcal/mole.

It is possible to construct molecules in which both Watson-Crick and Hoogsteen hydrogen bonding can be expressed simultaneously with aromatic stacking interactions. Specifically, the naphthalene spacer (fig. 5) provides a system that can chelate adenine derivatives. These show such extraordinary affinity for adenine that they extract adenine and even adenosine from aqueous solutions into chloroform[3]. To date, our most efficient binding of adenine occurs with a carbazole derivative; this spacer appears to match the shape of adenine with its complementary surfaces[5].

The natural use of hydrogen bonding and other such weak forces arises from their action over very short distances. High selectivity can be achieved at such levels of molecular intimacy, and changes in the acid-base characteristics of the hydrogen bonding patterns alters the selectivity in a profound way. For example, reduction of the imides leads to hydroxy lactams (fig. 6) which now show recognition of cytosine derivatives, and the selectivity over adenine has been reversed[9]. More recently, acylated amidines have also been prepared in this series; these exhibit high affinity for guanosine derivatives (fig. 6) and association constants are in the 10^3–10^4 range[19]. Parallel developments using macrocyclic molecules[7] have also led to selective binding of nucleic acid components.

Hydrogen-bonding patterns feature intrinsic energetics, and a recent perspicacious analysis by Jorgensen[10] has revealed subtle, secondary interactions that contribute to the stability of base pairs. For example, hydrogen-hydrogen repulsions and heavy atom-heavy atom repulsions destabilize base pairings such as shown in A (fig. 7), whereas in a seemingly similar pattern B, additional stabilization occurs. Using a series of imides and lactams, it has been possible to measure some of these effects in solution. For example, in the series shown (fig. 7) it is possible to expose the destabilizing effect of the spectator or nonparticipating oxygen on the base pairing event. Each such interaction destabilizes the nearby hydrogen bond by $\sim$ 0.4 kcal; as a consequence, lactams self-associate more than imides do[8].

In addition to the secondary effects, acid-base chemistry also contributes to the relative strengths of base pairing. For example, direct comparison of the imide and lactam (fig. 8) reveals the enhanced acidity of the imide can overcome the destabilizing influence of the remote carbonyl. Alternatively, the participation of this carbonyl in an unconventional hydrogen bond to H_2 or H_8 of the adenine nucleus may be responsible[8].

Self-replicating systems

Self-complementarity appears to be a fundamental requirement for a replicating system. This notion can be simply demonstrated through the *coupe du roi*[1]. A solid object such as the cube is cut into two (chiral) pieces. They are identical and self-complementary (fig. 9). The possibility that arises is that one of the pieces can act as a template for the construction of the other from its smaller sub-units.

At the molecular level, the quintessential structure is the double-stranded nucleic acid. A short, palindromic sequence (fig. 10) can be separated, and one single strand

Recognition of cytosine and guanine derivatives

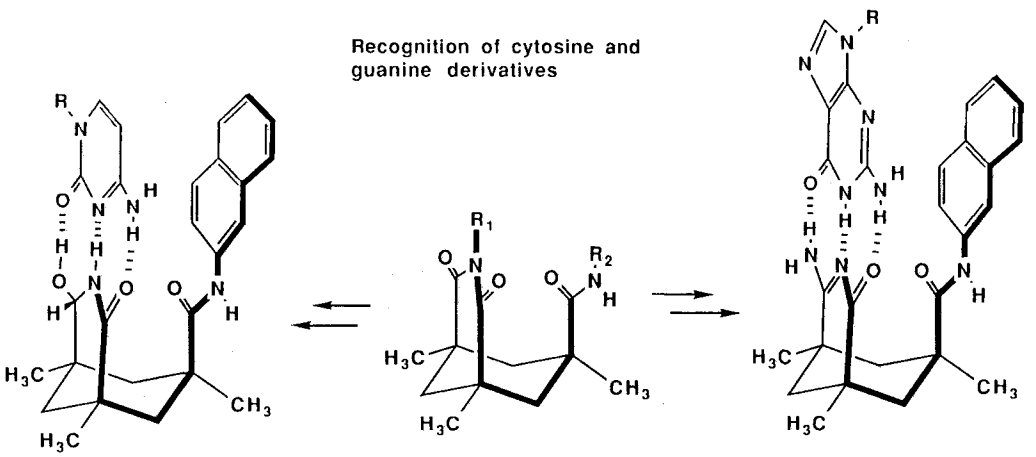

Figure 6

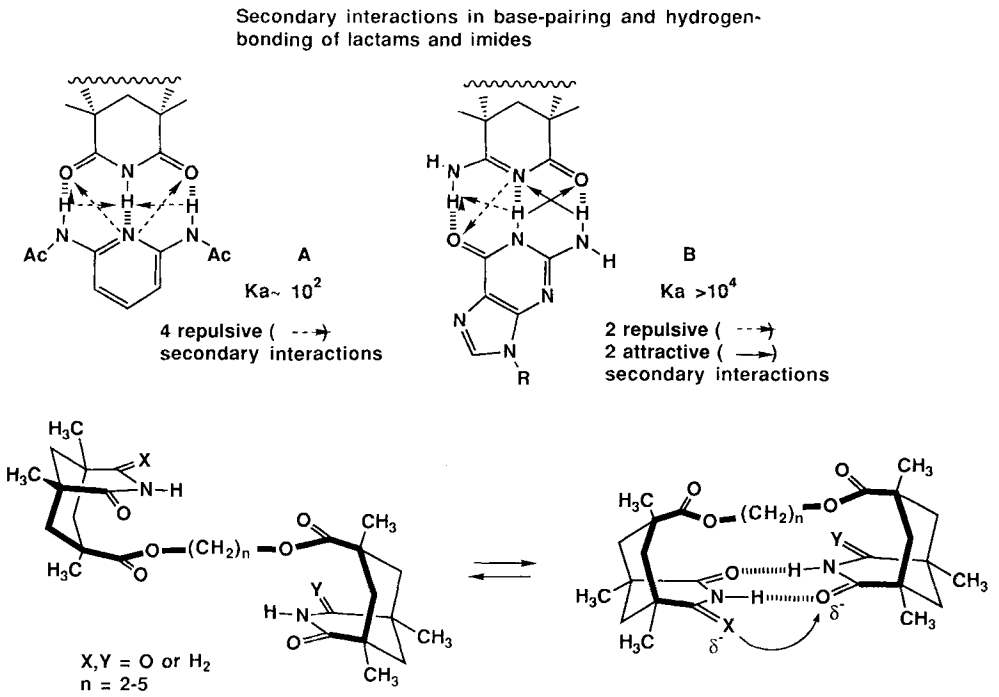

Figure 7

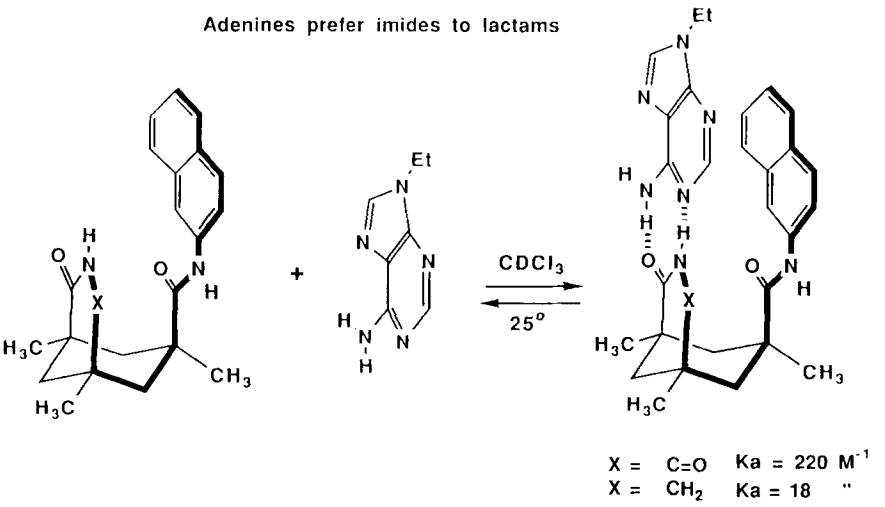

Figure 8

can act as a template for the assembly of the other. It is not surprising that this has been accomplished in phosphate chemistry, both enzymatically and nonenzymatically[21,24]. Indeed, it would be surprising if it couldn't, since this is the fundamental event of reproduction.

With the award of the 1989 Nobel Prizes for the discovery of ribozymes, more attention has been turned to the evolution of a protein world from an RNA world. In making the transition from nucleic acids to proteins, an enormous gap exists. Indeed, two molecules that are less similar would be hard to imagine. The RNA itself is a bizarre juxtaposition of domains. The bases ultimately originate from the gas phase condensation of simple molecules[13], while the phosphate arises from dissolving rocks. The carbohydrate emerges eventually from the third phase, an aqueous solution (fig. 11).

That nucleic acids taught proteins how to reproduce has posed one of the great questions since base pairing was discovered: what is the relationship between codon and the structure of the amino acid? What were the adaptor molecules that bridged the two worlds? Sober accounts are available[15], but this area has been so rife with speculation that it has acquired a rather unsavory reputation. A few key experiments are interpreted, in monographs

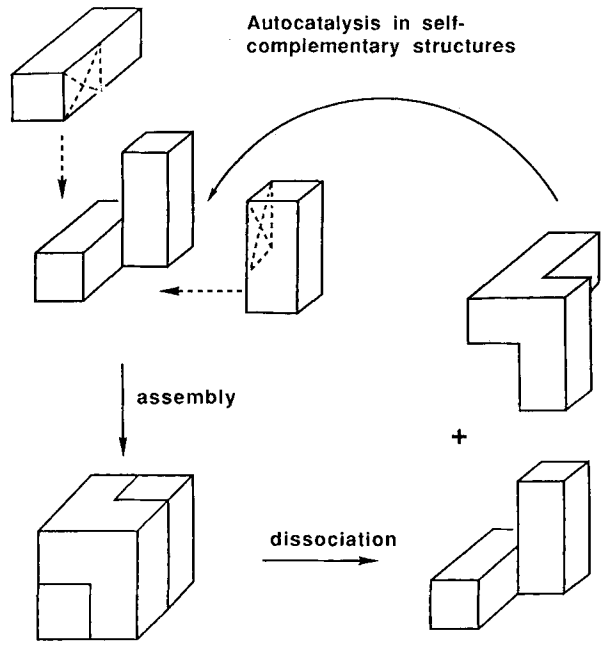

Figure 9

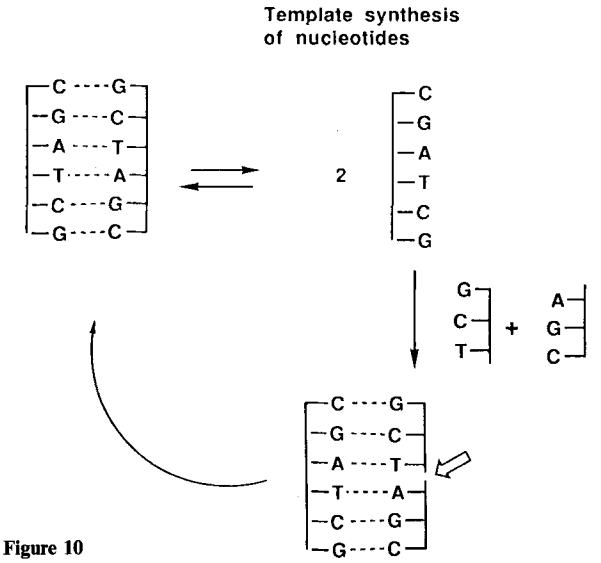

Figure 10

(often written at the authors' expense) to explain the mysteries.

We have devised a model system in which the information of the nucleic acid world is used to drive acyl transfer reactions to construct the amides of the protein world. By attaching an adenine *covalently* to a receptor for adenine, a compound is made that is, by definition, self-complementary. The point of departure is the actual linkage by which the two components are connected. Such structures can dimerize or oligomerize or, provided that the linkage is of suitable length and flexibility, it can fold shut in a self-satisfying sense (fig. 12).

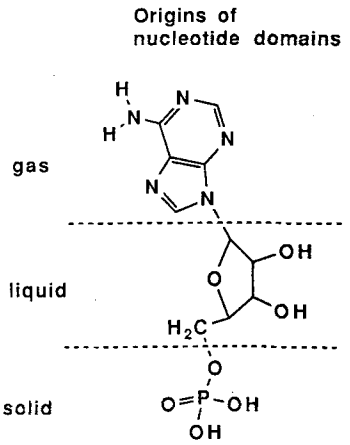

Figure 11

Our first *coupe du roi* structure was made quite inadvertently by attaching a quinoline derivative to the imides [2]. Such structures dimerized through complementary hydrogen-bonding and aryl-stacking interactions. The multipoint contacts of base pairing, however, provide greater structural advantages, and the control of Watson-Crick vs Hoogsteen bifurcated hydrogen bonds permitted the synthesis of a self-complementary system to take place in an orderly fashion. Specifically, the coupling of the phenyl derivative (fig. 13) with the adenosine derivative shown led to a structure with the minimal structural requirements of self-complementarity, and the first peculiarity was observed during the actual coupling reaction. In comparing the coupling rates of the NH vs the N-methyl compound, we found that the latter coupled some five or six times more slowly. This can hardly be attributed to a steric effect at a site so remote from the acyl transfer center. Rather, it is likely that the acyl transfer step itself proceeds by way of a base-paired intermediate. This was demonstrated to be the case by the use of competitive inhibitors. For example, 9-ethyl adenine slowed the coupling of the two components by tying up the electrophile in a nonproductive complex (molecular modeling indicated that the most likely way for the reaction to proceed was by way of Hoogsteen base-paired complex).

The final product of this reaction proved to be a great disappointment. It remained folded shut, that is, the short phenyl spacer and a *trans*-amide bond provide a length appropriate for intramolecular base pairing. NOE experiments indeed showed that this was the case, that no amount of dilution would pry two systems apart; the hydrogen-bonding edges and aryl-stacking surfaces were unavailable to act as templates for replication.

A longer spacer was clearly required. This was found in the 2,7-disubstituted naphthalene (fig. 14). This system showed the same peculiarities in the coupling reaction as its phenyl counterpart. Base pairing again precedes acyl transfer, most likely in the Watson-Crick mode. Here,

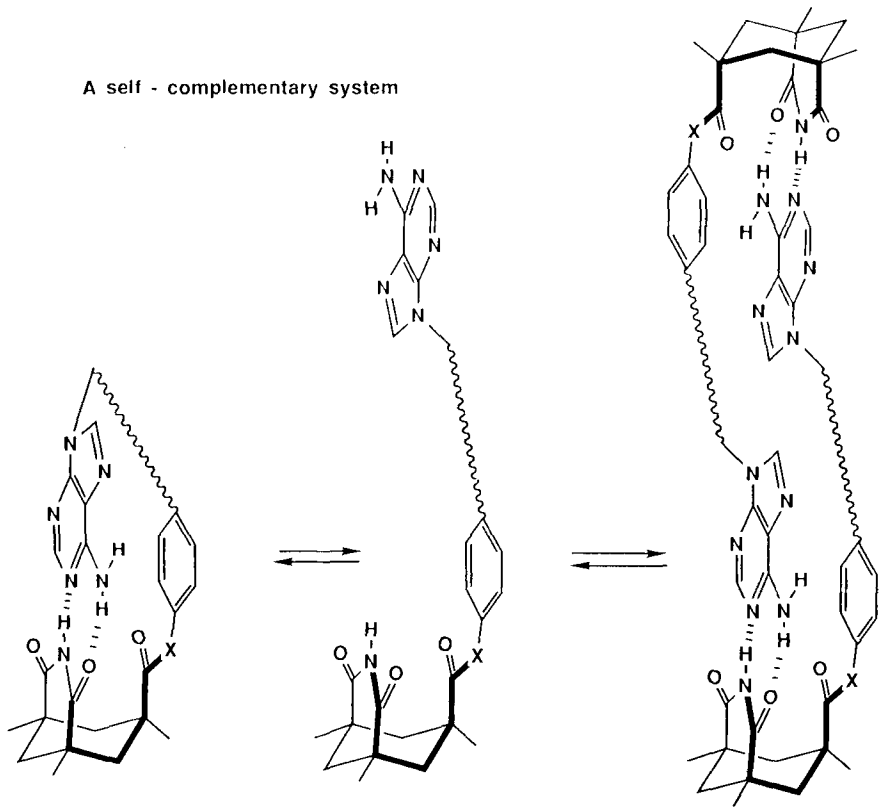

Figure 12

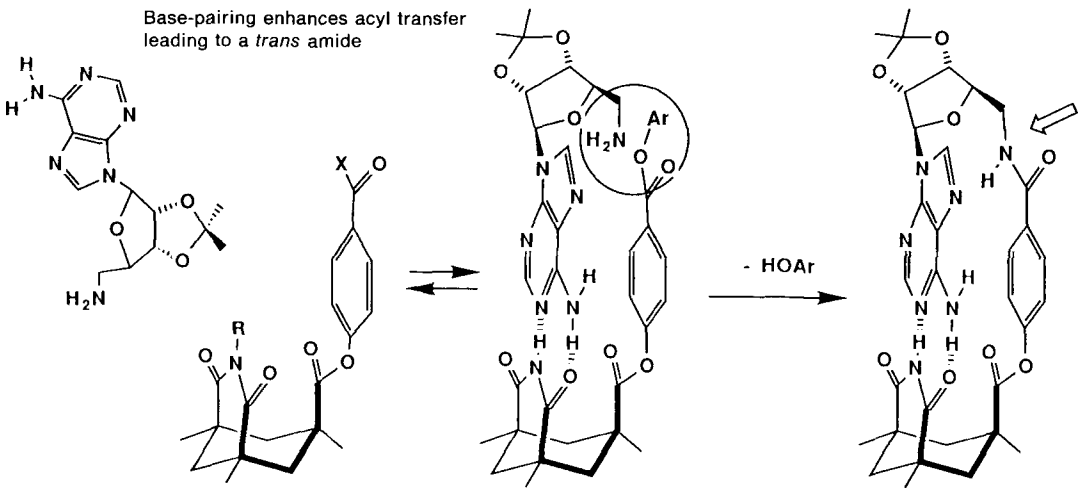

Figure 13

however, modeling suggests that the acyl transfer leads initially to a *cis*-amide bond. The feature provides a tension, or spring-loaded structure which jackknifes open to the more stable *trans*-amide form. The latter exposes the self-complementary sites, and indeed dimerization in the *coupe du roi* sense occurs. Fortunately this dimerization can be controlled; the bulges in the center of the molecule apparently destabilize the system somewhat and it can be pried apart by dilution.

The result is remarkable. The system does act as a template for its own construction, that is, it shows *autocatalysis*[20]. Addition of the product to coupling mixtures enhances the initial rates of formation. Apparently, a third form of coupling takes place, proceeding by way of a termolecular complex in which base pairing and aromatic stacking forces help position the electrophile and nucleophile for rapid coupling (fig. 15). From the association constants of the various equilibria, it is possible to

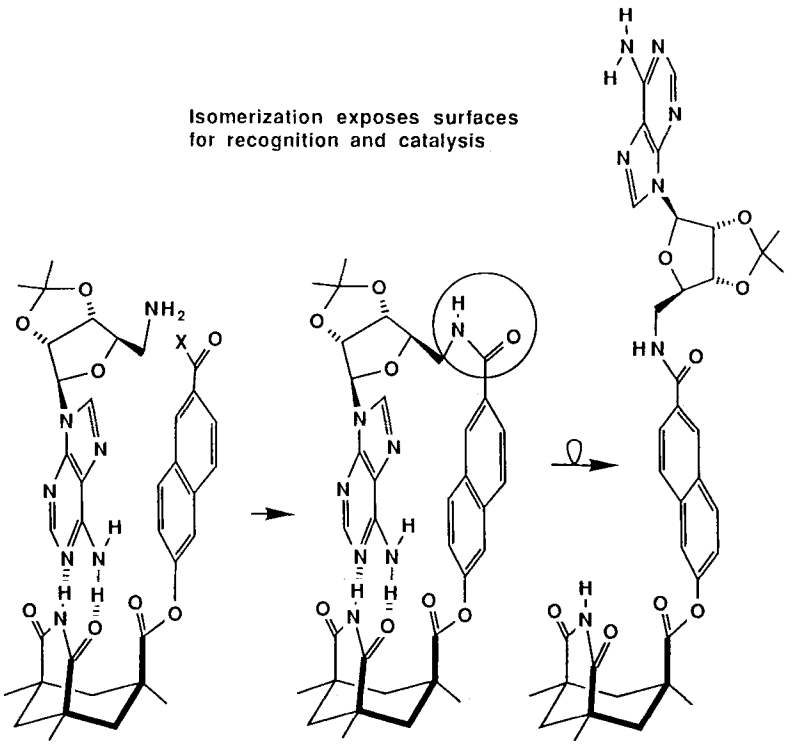

Figure 14

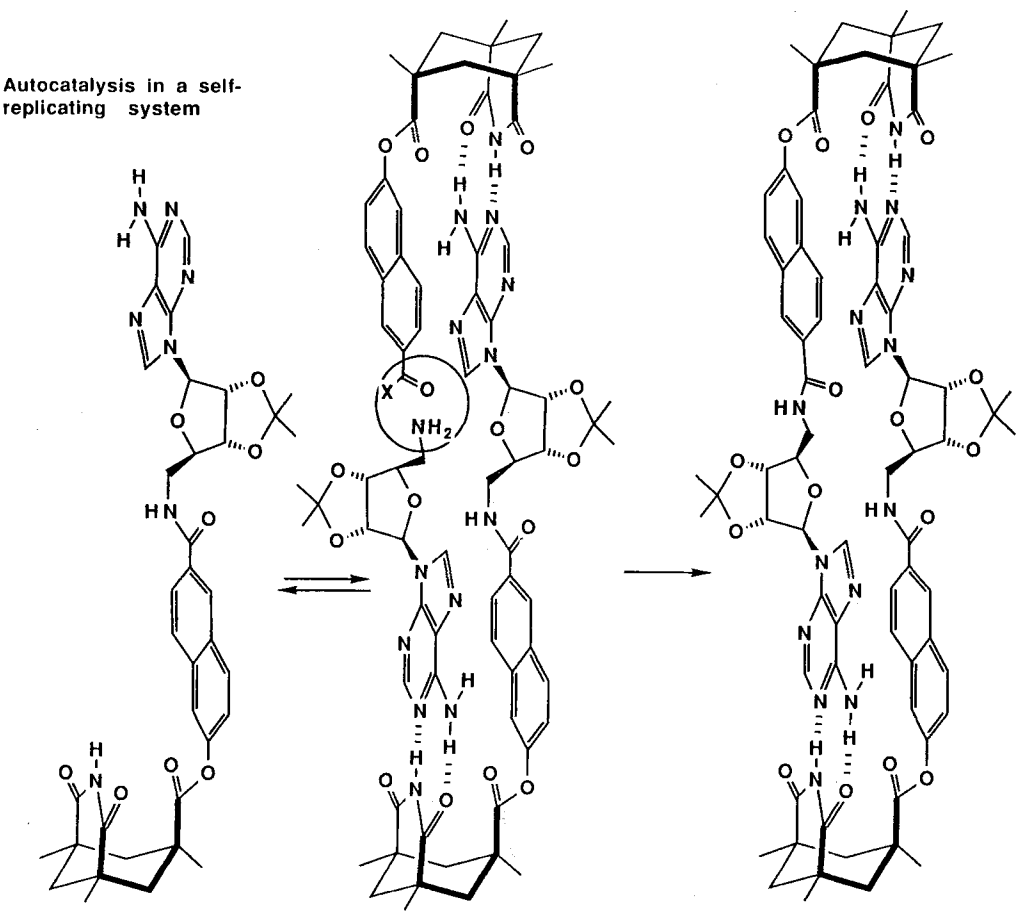

Figure 15

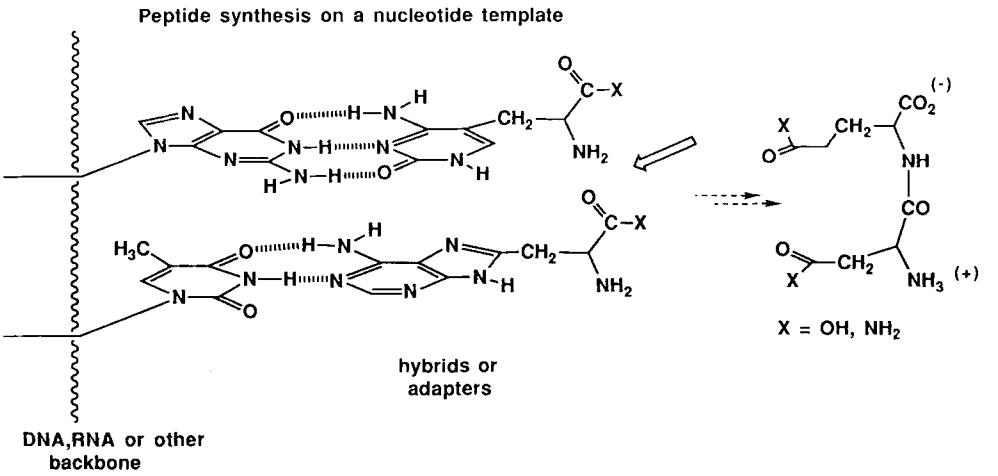

Figure 16

calculate that only about 2% of the mixture exists as the termolecular complex, yet the coupling rate increases by >30%. Thus the termolecular reaction must be very efficient, indeed.

The above has demonstrated that it is possible to use base-pairing events to drive and catalyze acyl transfer reactions, and we are working on more recognizable adaptor molecules to explore this phenomenon. For example, our current attention is focused on the use of hybrid structures, amino acids with purine and pyrimidine side chains (fig. 16). Subsequent hydrolysis of some of the condensation products would lead to recognizable amino acid side chains. These systems would provide a more direct link to how nucleic acids can be involved in the formation of longer chain peptides.

Why such an event was encouraged is a good question as well. The condensation that led to RNA's in the prebiotic world eventually gave some structures that showed catalytic activity of the nuclease sort. Such molecules would act as predators on other RNA's (or themselves, for that matter), and since they provided a self-replication capability, eventually all of the RNA's would be selected as predatory sequences. There is no evidence yet that such molecules show any protease activity, but at the same time or subsequently, the prebiotic condensations led to small peptides. Given the richness of functionality available in the side chains of these structures, some would show nuclease activity and some would show protease activity. Those featuring the latter would act as predators to digest exposed peptides.

We speculate that *protection* was the original motivating factor that drove the nucleotides and peptides together. That is, peptide nucleic acid complexes which protected both components from their respective predators would be advantageous for both, and would provide the motivation at least for nucleic acids to 'teach' selected peptides to reproduce. One form in which an advantageous contact may have taken place is shown in figure 17. Hydrogen bonding of the NH of a peptide of the 2'-OH

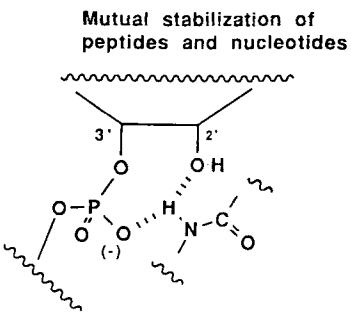

Figure 17

of a ribose could prevent the cyclization leading to intramolecular phosphate cleavage that is the inherent weakness of RNA. At the same time, such a contact could stabilize the peptide bond toward the approach of external nucleophiles. We are currently exploring this notion.

Acknowledgments. I am most grateful for the capable experimental assistance of my co-workers at the University of Pittsburgh and M.I.T. Their names appear in the original literature citations. I thank Prof. W. Jorgensen for helpful discussions and for sharing his results with us prior to their publication. Financial support was provided by the National Science Foundation through their Special Creativity Grant program.

1 Anet, F. A. L., Miura, S. S., Siegel, J., and Mislow, K., La Coupe du Roi and its relevance to stereochemistry. Combination of two homochiral molecules to give an achiral product. J. Am. chem. Soc. *105* (1983) 1419–1426.

2 Askew, B., Ballester, P., Buhr, C., Jeong, K. S., Jones, S., Parris, K., Williams, K., and Rebek, J. Jr, Molecular recognition with convergent functional groups VI: synthetic and structural studies with a model receptor for nucleic acid components. J. Am. chem. Soc. *111* (1989) 1082–1090.

3 Benzing, T., Tjivikua, T., Wolfe, J., and Rebek, J. Jr, Recognition and transport of adenine derivatives with synthetic receptors. Science *242* (1988) 266–267.

4 Chan, S. I., Schweitzer, M. P., Ts'o, P. O. P., and Helmkamp, G. K., Interaction and association of bases and nucleosides in aqueous solutions. III. A nuclear magnetic resonance study of the self-association of purine and 6-methylpurine. J. Am. chem. Soc. *86* (1964) 4182–4188.

5 Deslongchamps, G., Galan, A., de Mendoza, J., Rebek, J. Jr, A synthetic receptor for cyclic AMP. Angew. Chem. (1991) in press.

6 Ferguson, S. B., Seward, E. W., Diederich, F., Sanford, E. M., Chou, A., Inocencio-Szweda, P., and Knobler, C. B., Strong enthalpically driven complexation of neutral benzene guests in aqueous solution. J. org. Chem. *53* (1988) 5593–5595.
7 Hamilton, A. D., and Van Engen, D., Induced fit in synthetic receptors: Nucleotide base recognition by a 'Molecular Hinge'. J. Am. chem. Soc. *109* (1987) 5035–5036.
8 Jeong, K. S., Tjivikua, T., and Rebek, J. Jr, Relative hydrogen bonding affinities of imides and lactams. J. Am. chem. Soc. *112* (1990) 3215–3217. The unconventional hydrogen bond was brought to our attention by Prof. W. Jorgensen.
9 Jeong, K. S., and Rebek, J. Jr, Molecular recognition: Hydrogen bonding and aromatic stacking converge to bind cytosine derivatives. J. Am. chem. Soc. *110* (1988) 3327–3328.
10 Jorgensen, W., and Pranata, J., Importance of secondary interactions in triply hydrogen bonded complexes: Guanine-cytosine vs. uracil 2,6-diaminopyridine. J. Am. chem. Soc. *112* (1990) 2008–2010.
11 Kellis, J. T. Jr, Nyberg, K., Sali, D., and Fersht, A. R., Contribution of hydrophobic interactions to protein stability. Nature *333* (1988) 784–786.
12 Kemp, D. S., and Petrakis, S. K., Synthesis and conformational analysis of *cis*, *cis*-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid. J. org. Chem. *46* (1981) 5140–5143.
13 Ksander, G., Bold, G., Lattman, R., Lehman, C., Fruh, T., Xiang, Y.-B., Inomata, K., Buser, H. P., Schreiber, J., Zass, E., and Eschenmoser, A., Chemie der α-Amino Nitrile. Helv. Chim. Acta *70* (1987) 1115–1172.
14 Kyogoku, Y., Lord, R. G., and Rich, A., The effect of substituents on the hydrogen bonding of adenine and uracil derivatives. Proc. natl Acad. Sci. USA *57* (1967) 250–257.
15 Orgel, L. E., Evolution of the genetic apparatus: A review. Cold Spring Harbor Sym. Quant. Biol. *52* (1987) 9–16.
16 Rebek, J. Jr, Askew, B., Ballester, P., Buhr, C., Costero, A., Jones, S., and Williams, K., Molecular recognition: Watson-Crick, Hoogsteen and bifurcated hydrogen bonding in a model for adenine recognition. J. Am. chem. Soc. *109* (1987) 6866–6867.
17 Rebek, J. Jr, Askew, B., Ballester, P., Buhr, C., Jones, S., Nemeth, D., and Williams, K., Molecular recognition: Hydrogen bonding and stacking interactions stabilize a model for nucleic acid structure. J. Am. chem. Soc. *109* (1987) 5033–5035.
18 Saenger, W., Principles of Nucleic Acid Structure, Ch. 6. Springer-Verlag, New York 1984.
19 Schroeder, J., Park, T.-K., and Rebek, J. Jr, Convergent functional groups XI. Selective recognition of guanosine derivatives Tetrahedron *47* (1991) 2507–2518.
20 Tjivikua, T., Ballester, P., and Rebek, J. Jr, A self-replicating system. J. Am. chem. Soc. *112* (1990) 1249–1250.
21 von Kiedrowski, G., Woltzka, B., and Helbing, B., Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3′-5′ pyrophosphate linkage. Angew. Chem. Int. Ed. Eng. *28* (1989) 1235–1237.
22 Williams, K., Askew, B., Ballester, P., Buhr, C., Jeong, K.-S., Jones, S., and Rebek, J. Jr, Molecular recognition with convergent functional groups VII: Energetics of adenine binding with model receptors. J. Am. chem. Soc. *111* (1989) 1090–1094.
23 Williams, N. G., Williams, L. D., and Shaw, B. R., Dimers, trimers, and tetramers of cytosine with guanine. J. Am. chem. Soc. *111* (1989) 7205–7209 and references therein.
24 Zielinski, W. S., and Orgel, L., Autocatalytic synthesis of a tetranucleotide analogue. Nature *328* (1987) 346–347.

0014-4754/91/11-12/1096-09$1.50 + 0.20/0

A comparison of pyridoxal 5′-phosphate dependent decarboxylase and transaminase enzymes at a molecular level

D. M. Smith, N. R. Thomas and D. Gani*

*Department of Chemistry, The Purdie Building, The University, St. Andrews, Fife KY16 9ST (Scotland)***

Abstract. Pyridoxal 5′-phosphate is a coenzyme for a number of enzymes which catalyse reactions at C^{α} of amino acid substrates including transaminases, decarboxylases and serine hydroxymethyltransferase. Using the X-ray coordinates for a transaminase, aspartate aminotransferase, and the results of stereochemical and mechanistic studies for decarboxylases and serine hydroxymethyltransferase, an active-site structure for the decarboxylase group is constructed. The structure of the active-site is further refined through active-site pyridoxyllysine peptide sequence comparison and a 3-D catalytic mechanism for the L-α-amino acid decarboxylases is proposed. The chemistry of serine hydroxymethyltransferase is re-examined in the light of the proposed decarboxylase mechanism.
Key words. Transaminase; decarboxylase; serine hydroxymethyltransferase; pyridoxal 5′-phosphate; enzyme mechanism; stereochemistry; kinetics.

Introduction

Pyridoxal 5′-phosphate (PLP) dependent enzymes catalyse a myriad of different types of chemical reaction in amino acid metabolism. Examples include α-amino acid-keto acid and ω-amine-aldehyde transamination; α-, and β-amino acid decarboxylation; α-amino acid racemisation; β-hydroxy amino acid retro-aldol cleavage; α-amino acid deamination; and elimination and replacement reactions at C^{β} and C^{γ}. Some examples are shown in scheme 1.

Of the PLP-enzyme catalyzed reactions, the α-amino acid transaminase and α-amino acid decarboxylase groups are the largest. In each group there are dozens of different enzymes which catalyse the conversion of specific substrates to specific products. These have different names and Enzyme Catalogue[24] (EC) numbers, for example, glutamate and histidine decarboxylase. There are also several quite distinct characterised enzymes, from different species, which catalyse the same specific reaction.