Account

J. Mex. Chem. Soc. **2009**, *53(1),* 23-33

ISSN 1870-249X

A Search for Structural Alternatives of RNA

Ramanarayanan Krishnamurthy

The Scripps Research Institute. 10550 North Torrey Pines Road. La Jolla, CA 92037 USA. rkrishna@scripps.edu

Received February 27, 2009; accepted March 25, 2009

Abstract. This account describes a search for potentially primordial informational oligomers; the work is the direct outcome of the research program that was initiated by the Eschenmoser group —at ETH Zürich about 20 years ago and was continued at The Scripps Research Institute since 1996— in order to understand the chemical basis on which nucleic acids were chosen by nature as the molecular foundation of its genetic apparatus. The investigation began with the study of the base-pairing properties of structural alternatives of nucleic acids —constructed from different sugars (hexo- & pento-pyranoses and tetrofuranose) retaining the canonical nucleobases and phosphates. The outcome from these studies led to the conclusion that Watson-Crick type base-pairing is not unique to RNA/DNA, and that it can be compatible with a wide variety of backbone edifice. This provided the motivation to map the landscape of potentially primordial informational oligomer systems that may contain backbones, recognition elements and linker groups structurally quite different from those known so far. The oligomer systems chosen for study are, conceptually, deemed to be (a) *potentially primordial* (based on the nature of the starting materials and reaction conditions considered to be prebiotically realistic) and (b) *informational* (based on their ability to adopt a repetitive conformation such that the information encoded by the recognition elements can be transmitted intermolecularly). Though such studies suggest the possibility of finding informational systems that could lay claim as functional ancestors of RNA —they are more likely to generate results that provide the opportunity to assess the structural and functional uniqueness of nature's choice.

The experimental investigation described here deals with the base-pairing properties of oligomer systems derived from 2,4-disubsituted -triazines, -5-aminopyrimdines and -6-carboxy pyrimidines as recognition elements that are tagged to oligo-dipeptide backbones via different linker groups. The results from the inter- and intra-system cross-pairing studies reveal that there is, on first approximation, a direct correlation between the magnitude of the difference in Δ*pKa* of the recognition elements and their base-pairing strength —*smaller the* Δ*pka between the base-pairing partners, weaker is the base-pairing strength* (in aqueous medium at near neutral pH). These results exemplify the inherent singularity of the canonical nucleobases— their ability to remain un-ionized under physiological conditions based on their constitution – emphasizing the relationship between their physicochemical properties and their functional competence in connection with their role in informational base-pairing.
Key words. Nucleic acids, structural alternatives, RNA, DNA, oligonucleotides, oligopeptides, pairing properties, canonical nucleobases, pKa, ionization.

Resumen. Este recuento describe la búsqueda de oligómeros informacionales potencialmente primordiales, resultado del programa de investigación iniciado por el grupo de Eschenmoser —en el ETH Zürich hace veinte años y continuado en el Instituto de Investigación Scripps desde 1996— dirigido al entendimiento de las bases químicas por las cuales los ácidos nucleicos fueron escogidos por la naturaleza como el fundamento molecular del sistema genético. La investigación empezó con el estudio de las propiedades de apareamiento de las bases de alternativas estructurales de ácidos nucleicos —construidos a partir de diferentes azúcares (hexo- & pento- piranosas y tetrofuranosas) reteniendo las nucleobases y los fosfatos. Los resultados condujeron a la conclusión de que el apareamiento tipo Watson-Crick de las bases no es exclusivo para ARN/ADN, y que este puede ser compatible con una amplia variedad de esqueletos. Estos resultados proporcionaron la motivación para explorar sistemas oligoméricos informacionales potencialmente primordiales que pudieran contener esqueletos, elementos de reconocimiento y grupos conectores diferentes a los conocidos hasta ahora. Los sistemas oligoméricos escogidos para estudio son considerados conceptualmente como (a) *potencialmente primordiales* (por la naturaleza de los materiales de partida y las condiciones de reacción, considerados realísticamente como prebióticos) y (b) *informacionales* (basado en su habilidad de adoptar una conformación repetitiva tal que la información codificada por los elementos de reconocimiento pueda ser transmitida intermolecularmente). Aunque estos estudios sugieren la posibilidad de encontrar sistemas informacionales que pudieran considerarse como ancestros funcionales del ARN, mas bien generan resultados que proporcionan la oportunidad de evaluar la singularidad funcional de la selección de la naturaleza.

La investigación experimental descrita aquí se refiere a las propiedades de apareamiento de las bases de sistemas de oligómeros derivados de triazinas 2,4-sustituidas, -5-aminopirimidinas y -6-carboxi-pirimidinas como elementos de reconocimiento unidos a esqueletos de oligopéptidos mediante diferentes grupos conectores. Los resultados de los estudios del apareamiento cruzado inter- e intra- sistema revelan que existe, en una primera aproximación, una correlación directa entre la magnitud de la diferencia en Δ*pKa* de los elementos de reconocimiento y la fuerza del apareamiento de las bases. *A menor* Δ*pKa entre los elementos de apareamiento, menor la fuerza de apareamiento de las bases* (en medio acuoso a pH cercano al neutral). Estos resultados ejemplifican la singularidad inherente de las nucleobases canónicas —su habilidad de permanecer no ionizadas en condiciones fisiológicas de acuerdo a su constitución— enfatizando la relación entre sus propiedades fisicoquímicas y su competencia funcional en relación con su papel en el apareamiento informacional de las bases.
Palabras clave. Ácidos nucleicos, alternativas estructurales, ARN, ADN, oligonucleótidos, oligopéptidos, propiedades de apareamiento, nucleobases canónicas, pKa, ionización.

*Editorial note***.** This account is the written version of the lecture presented by Prof. Ramanarayanan Krishnamurthy at the Symposium *Origins of Life in the Earth*, organized by El Colegio Nacional and Sociedad Química de México, held at El Colegio Nacional, Donceles 104, Centro Histórico, México, D.F., on March 18, 2009.

Introduction

The question, "why nature chose the structure type of ribofuranosyl nucleic acids (RNA), rather than some other family of molecular structures, as the molecular basis of life's genetic systems" [1] has led to an extensive research undertaking by the Eschenmoser group via an approach termed "Chemical Etiology of Nucleic Acid Stucture" [1]. Understanding the chemical causes and origins as to why RNA has been made-up of it components —ribofuranose, the canonical nucleobases and phosphate— has been at the core of the research project initiated by Eschenmoser [2,3] since the late 1980s at the ETH in Zürich (and since 1996 together with the author at the Scripps Research Institute in La Jolla CA).

Alternatives from the structural neighborhood of RNA. The inquiry began with questioning the sugar component of RNA: (a) why a pentose and not a hexose? (b) why ribose and not any other pentose?, and (c) why ribo-furanose and not ribo-pyranose? Such questions were motivated by observations that potentially natural reaction pathways starting from simple precursor phosphorylated aldehydes (such as glycolaldehye & glyceraldehyde phosphates) while leading to phosphorylated ribose, also led to the formation of other sugars such as hexose phosphates and other aldo-pentose phosphates. Assuming that the same type of chemistry that led to the formation of RNA from ribose were operative, it would be possible to imagine a scenario where there could be oligonucleotides derived from alternative sugar phosphate building blocks.

The Eschenmoser group has carried out comprehensive experimental investigations of oligonucleotides derived from diverse sugar phosphate alternatives taken from the structural neighborhood of RNA (Fig. 1) ranging from hexo-pyranoses to pento-pyranoses to tetro-furanoses [4-6]. While the β-hexopyranosyl-(6'→4') oligonucleotide analogs [4] of RNA derived from hexose sugars allose, altrose and glucose display Watson-Crick base-pairing far inferior to that of RNA with respect to both pairing strength and pairing mode specificity, the whole family of diastereomeric pentopyranosyl-(4'→2') oligonucleotide analogs [5] of RNA including pyranosyl-RNA ("p-RNA") show Watson-Crick base-pairing within the family that is *uniformly stronger* than RNA itself. All of the oligonucleotide analogs derived from the hexopyranosyl-(6'→4') and pentopyranosyl-(4'→2') sugar-phosphate backbones do not cross-pair with the natural systems RNA or DNA. In the tetrose series, the correspondingly derived α-threofuranosyl oligonucleotide (TNA), was shown not only to be an informational Watson-Crick base-pairing

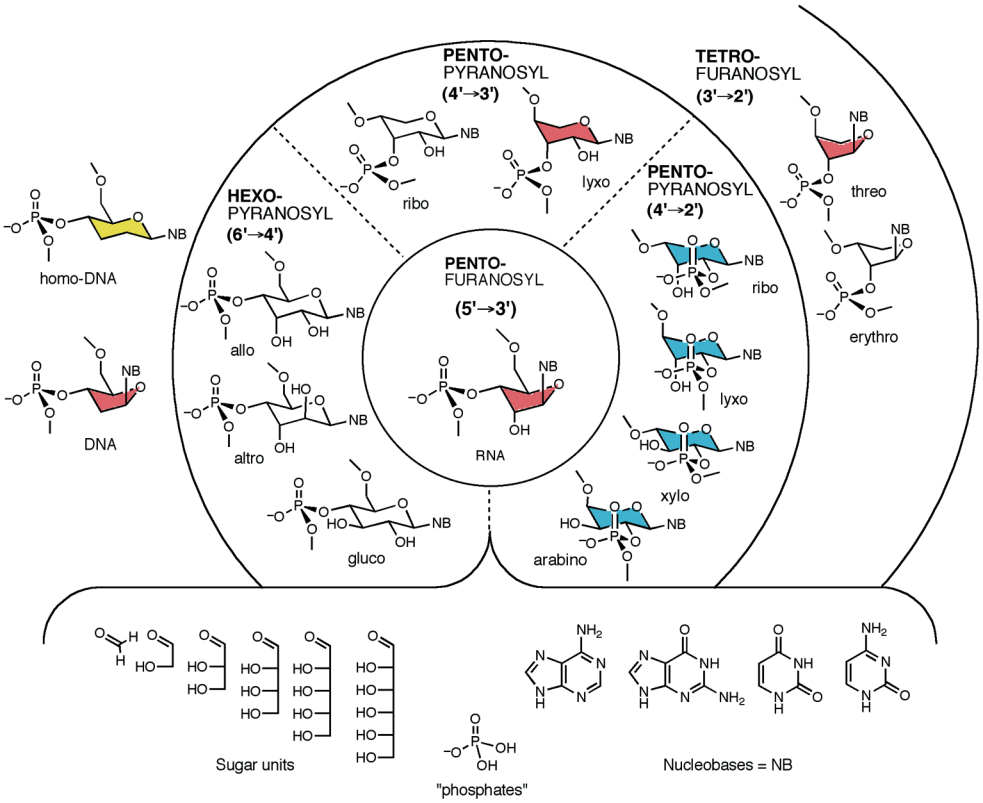

Fig. 1. The various oligonucleotides composed of alternative sugar phosphate backbones, chosen from the structural neighborhood of RNA. These alternatives are considered to be accessible starting from sugar unit building blocks (HCHO), phosphates and the canonical nucleobases by the same type of potentially natural chemistry that is presumed to have led to the formation of RNA from its corresponding building blocks. Homo-DNA (shaded yellow), which was studied as a model system, is not considered to be a potentially natural system according to this criterion. Pairing-systems (in color) exhibit inter-system base-pairing with systems of the same color.

system but also to have the capability of Watson-Crick cross-pairing with RNA and DNA [6].

The central conclusions from the above studies are:

- Watson-Crick base-pairing is not *unique* to the ribofuranoid oligonucleotide system.
- Watson-Crick base-pairing is not *specific* to the double-helix type of conformation characteristic of RNA (& DNA).
- Nature has not chosen RNA (& DNA) based on the criterion of *maximizing* base-pairing strength; rather, RNA seems to represent the *optimum* with respect to base-pairing capacity.
- A *six-bond* periodicity (as occurring in RNA & DNA) is not a constitutional pre-requisite for Watson-Crick base-pairing to be operative.
- Watson-Crick base-pairing can be *compatible* with a still largely unexplored structural landscape of informational oligomer systems that may contain backbones, recognition elements and linker groups structurally quite *different* from those known so far.

That the spectrum of potentially natural structural alternatives could be expanded beyond the environs of sugar-alternatives (a conclusion that is reinforced by the existence of various backbone modified oligomers, *e.g.* PNA [7], in the antisense research field [8]) led Eschenmoser to consider a broad variety of oligomer systems that could be deemed as potentially primordial [9]. This entailed experimentally mapping the landscape of potential informational oligomer systems with any type of backbone, any type of recognition element —as long as the structures are derivable from starting materials, and under reaction conditions, that are prebiotically relevant. A structurally unconstrained experimental exploration of this landscape with a strict focus on such potentially primordial members represents a continuation of the research pursued so far but in a much broader context, as the target criterion would *no longer be the search for function among structural RNA-alternatives of a generational complexity similar to that of RNA*, *but rather a search for generational simplicity among (functional) oligomers*, a simplicity that justifies a given system to be classified as potentially primordial. Though such expanded studies would be close to research in prebiotic chemistry and be driven by the possibility of finding informational systems that could have been functional ancestors of RNA— the outcomes of these studies are more likely to shed light on the chemical reasons as to why the constituents of RNA are 'special' and represent an 'optimum' with respect to its structure and function, and thereby create a perspective for the assessment of the degree and structural and functional uniqueness of nature's choice.

Expanding the scope in the search for alternatives to RNA. We have initiated a systematic experimental search for potentially primordial informational oligomers. The structures of these alternative oligomers are deduced, conceptually, by combining two strategies: (a) a *forward-synthetic analysis* based on organic source compounds that may have been available on early earth, and (b) a *qualitative conformational analysis* to assess a given oligomer's base-pairing capability.

(a) Forward-synthetic analysis: In this approach, we proceed in a methodical manner, starting from a range of elementary source materials (*e.g.* HCN, NH_3, HCHO, NH_2CN, etc.) and organic compounds postulated to have been available on early earth. By obeying rather strictly defined constraints of prebiotic chemistry, relying on the basic rules of organic chemistry we arrive at a collection of alternative, potentially primordial, nucleobases and backbone linker units. From a library of these "structural units", we then identify building blocks of potentially informational oligomer systems, to be synthesized and studied. In spite of the enduring uncertainties regarding the nature of such starting materials and environmental conditions of the prebiotic world, this approach is a reasonably practical one to reach an experimentally based conclusion on the structural diversity of potentially primordial genetic systems. (b) Qualitative conformational analysis: A qualitative conformational analysis to gauge the potential of a given oligonucleotide structure to base-pair was developed by Eschenmoser [2,10] and has been used, with a large measure of success, to predict the presence or absence of base-pairing conformations of oligonucleotide structures based on alternative sugar-phosphate backbones. The alternative oligomer candidates derived from the *forward-synthetic analysis* approach are subjected to this qualitative conformational analysis to gauge their potential to function as informational base-pairing systems; those systems that fulfill this criteria are then chosen to be studied further.

Mapping the landscape of potentially primordial informational oligomers: The search for the building blocks of potentially primordial informational oligomers, keeping in mind the generational caveats posed above, is split into two categories: (a) alternatives to the canonical nucleobases and (b) alternatives to the backbone linkers.

In choosing alternative recognition elements, potentially natural heterocycles that could be conceivably formed as alternatives to the natural nucleobases, we were guided by two important aspects —both of which are reflective of the strengths and weaknesses of the canonical nucleobases: (a) the potential of these alternative heterocycles to function as recognition elements (*i.e.* to act as a hydrogen bond forming base-pairing partner), and (b) their chemical potential for attaching themselves to a backbone building unit. This latter aspect is brought forcefully into play by the canonical nucleobase's demonstrated inability to attach themselves to the sugar building block under potentially natural conditions in an efficient manner ('nucleosidation problem') [11].

The first alternative canonical heterocycle chosen for study, belonged to the triazine family (2,4-diamino-triazine derivatives and their oxo-analogs, Fig. 2c). We were led to consider triazines as recognition elements while studying another class of heterocycles, allopurines [12] —which was

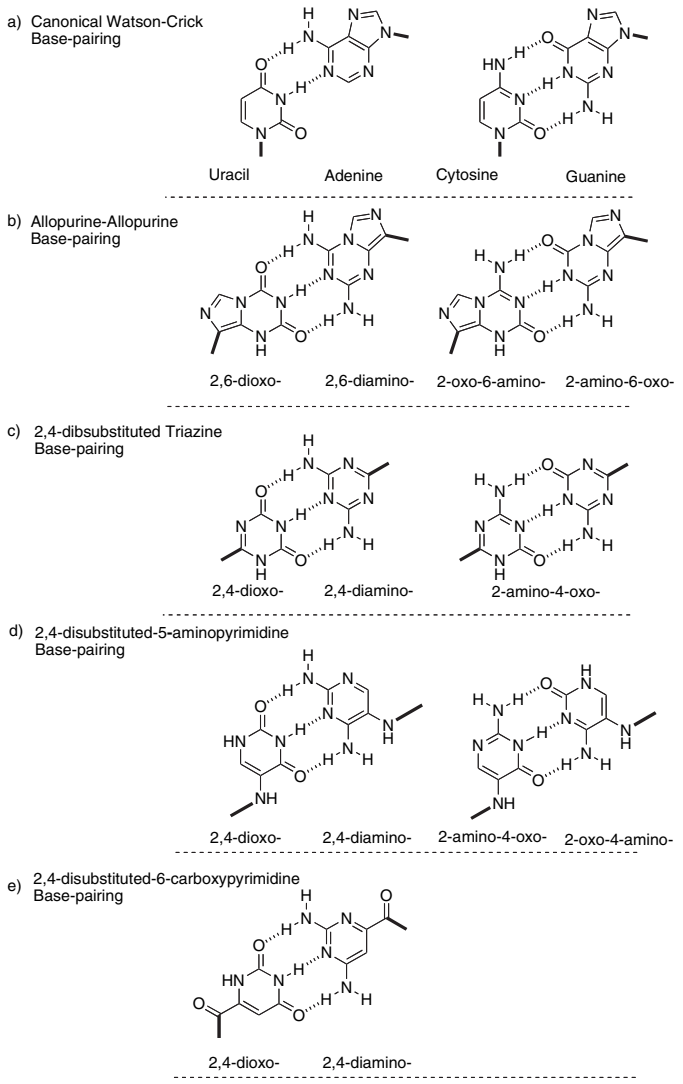

Fig. 2. The juxtaposition of (a) Watson-Crick type base-pairing system with those derived from potentially natural alternative heterocycles: (b) allopurines base-pairs; (c) triazine-derived base-pairs (the base-pair on the right hand side is composed of two tautomers of the same base); (d) 5-aminopyrimidine derived base-pairs; (e) 6-carboxypyrimidine derived base-pairs. The base-pairing – backbone axis (indicated by a bold line) in the alternative heterocycles is different from that in the natural (Watson-Crick type) base-pairing system.

derived by a forward synthetic analysis in the context of the nucleosidation problem. Triazines can be produced from the combination of biguanide and glycine, two relatively simple starting materials [13]; also, triazine heterocycles have been shown to easily form from a variety of carboxyl derivatives [13]. The 2,4-disubstituted triazine-rings retain the potential to function as informational base-pairs, but through a type of hydrogen bond arrangement (centro-symmetric) that differs from the canonical Watson-Crick type with its pairing axis being parallel to the nucleosidic bond (Fig. 2c).

The second family of heterocycle that is an etiologically appealing alternative to natural nucleobases is the 2,4-disubstituted 5-aminopyrimidine and its derivatives (Fig. 2d). The 5-aminopyrimidines have been shown to form from ammonium cyanide solutions, along with a wide variety of pyrimidines (and purines) [14]. The 5-amino group in this family of heterocycles, is expected to be the most nucleophilic among all the amino groups, and, therefore would be expected to form amide bonds when reacted with various carboxyl groups. Analogous to the triazine family of heterocycles, the 2,4-disubstituted 5-aminopyrimidines exhibit the potential to function as informational base-pairs.

The third family of heterocycles that we considered is the one derived from 6-carboxypyrimidines (Fig. 2e). The famous member of this family is orotic acid; not only is orotic acid relevant from a prebiotic view point (its ease of formation from hydrolysis of NH_4CN solutions parallels that of the canonical purines and other pyrimidines) [14c,15], but also from the standpoint of contemporary biology, as a precursor in the *de-novo* synthesis of uridine. The 6-carboxyl group in this series of heterocycles could be used as an attachment point to a backbone assembly unit while the 2,4-subsitutents allow for it to serve as informational base-pair.

Alternative backbone linkers. The alternative heterocycles considered above call for a carboxyl unit (for the triazines and 5-aminopyrimidines) or an amino unit (for the 6-carboxypyrimidine derivatives) as point of attachment at the backbone. Out of the possible carboxyl- and amine- containing-source-compounds, an attractive and potentially natural source is α-amino acids. The carboxylic group and the amino group of an amino acid lend themselves not only to oligomerization to form peptides, but can also act as points of attaching a recognition element via the formation of an amide bond. Oligoamides and oligopeptides have been shown to function as skeletons of informational oligomers [16].

For our study, we selected dicarboxylic amino acids such as aspartic and glutamic acid, where the β- and γ- carboxylic acid groups can be converted to a recognition element (triazines) or tagged with 5-amino group of the 5-aminopyrimidines, while the α-carboxylic group can lend itself to be part of the oligomeric backbone by forming peptide linkages (Fig. 3). In the case of 6-carboxypyrimidine derivatives, we chose 3-aminoalanine [17] as the building unit; the β-amino group can be tagged with the 6-carboxyl group of orotic acid derivatives, while the α-amino group can be part of the oligopeptide backbone. The role of an oligopeptide as a backbone for informational oligomeric systems has been hampered by its insolubility in water. To alleviate this problem, strategic placement of dicarboxylic acids at tactical (alternating) linking positions was adopted; the free β- or γ-carboxylic acid groups can contribute to the water solubility of the resultant oligopeptides. Thus, an assortment of oligopeptide backbone linker units tagged with 2,4-disubstituted- triazines, 5-aminopyrimidines, and 6-carboxypyrimidines (in place of conventional nucleobases) was envisaged (Fig. 3).

Qualitative conformational analysis. The strategy developed for assessing (idealized) conformational preferences of the

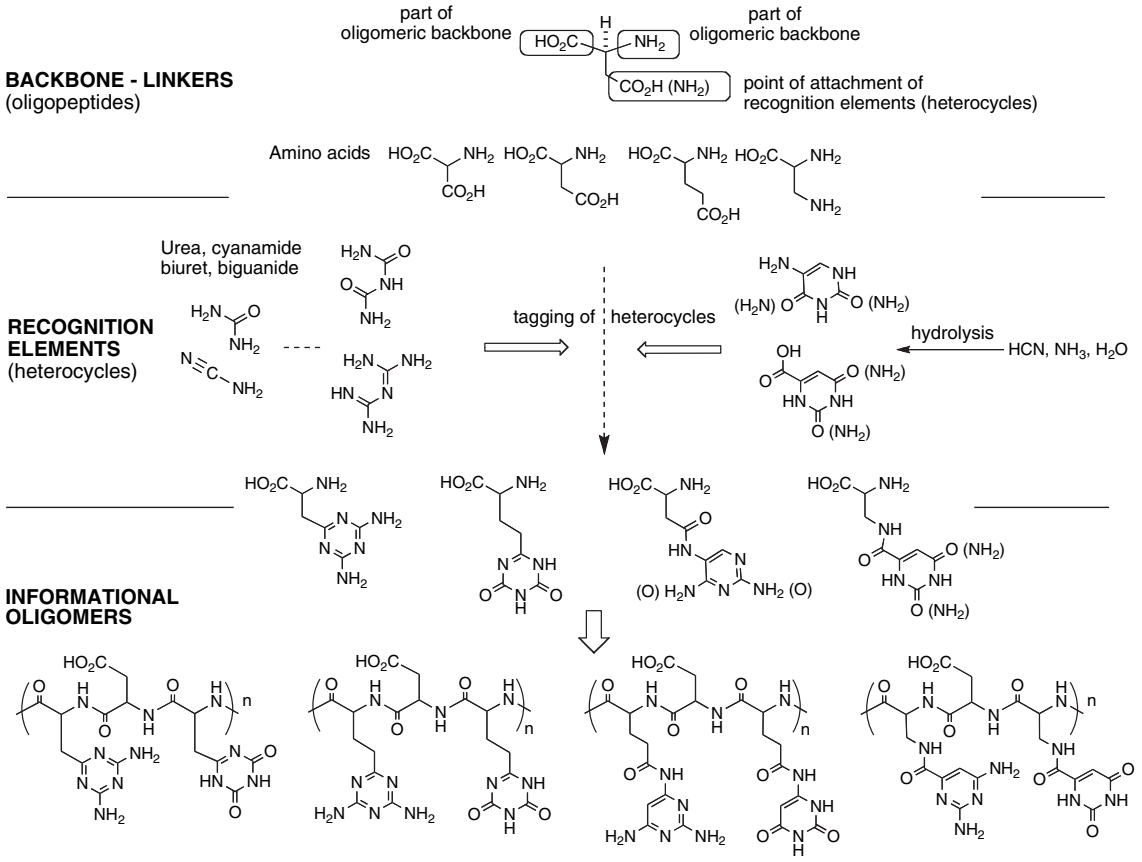

Fig. 3: An outline of the type of "forward-synthetic analysis" that is used in proposing potentially natural alternative informational oligomeric systems. Shown here is the one used in arriving at oligopeptides tagged with non-canonical recognition elements. These alternative systems could be conceived to form from natural source molecules (amino acids, cyanide and urea derivatives) under potentially prebiotic constraints via potentially primordial reactions involving the same type of elementary condensation chemistry (formation of amide bonds) at various levels.

cyclic variants derived from the structural neighborhood of RNA was applied, with modifications, to the acyclic oligomeric candidates [10]. For *e.g.* the sugar-phosphate backbone outline of A-form of DNA (RNA) was extracted, leaving out the cyclic part. This resulted in an acyclic-scaffold onto which recognition elements could be positioned at regular intervals (Fig. 4a); this framework was used as a starting point to assist in analyzing conformational preferences of assorted acyclic oligopeptide systems conceived above, and to gauge their prospect for adopting least strained conformations that would allow them to function as informational systems, via constitutionally repeating monomer base-pair units. Such an analysis reveals that the oligopeptide backbone, which is the common scaffold, is able to approximate the phosphodiester-backbone outline of A-form of DNA (RNA). Moreover, this approximation requires the constitutionally repeating monomer units of the oligopeptide to be dipeptide building blocks, so that there is conformity between the placements of recognition units at regular repeating intervals and those present in the A-form of DNA (RNA). Such an outcome fits in neatly with the strategy of having an amino acid linker unit at alternative positions to overcome the solubility problem associated with oligopeptides (Fig. 3). The orientation of hydrogen bonds of the recognition elements, relative to the backbone axis, can be manipulated by varying not only the length of the linker arm, but also by varying the point of attachment of linker on the heterocycle. Thus, the above qualitative conformational reasoning predicts that oligopeptides tagged with the three different families of heterocycles do harbor, not only the capacity to cross-pair with RNA and DNA, but also the capability to base-pair among themselves (Fig. 4b).

Results

Triazine and 5-aminopyrimidine tagged oligopeptides. The synthesis of (L)-oligopeptides tagged with 2,4-disubstituted-triazines and 5-aminopyrimidines were accomplished using standard synthetic protocols. Their cross-pairing properties with corresponding base-pairing counterparts in DNA and RNA, and their propensity for base-pairing among themselves, were studied [18].

As expected, homobasic sequences containing 2,4-diamino-triazine tagged oligopeptides, base-paired strongly with

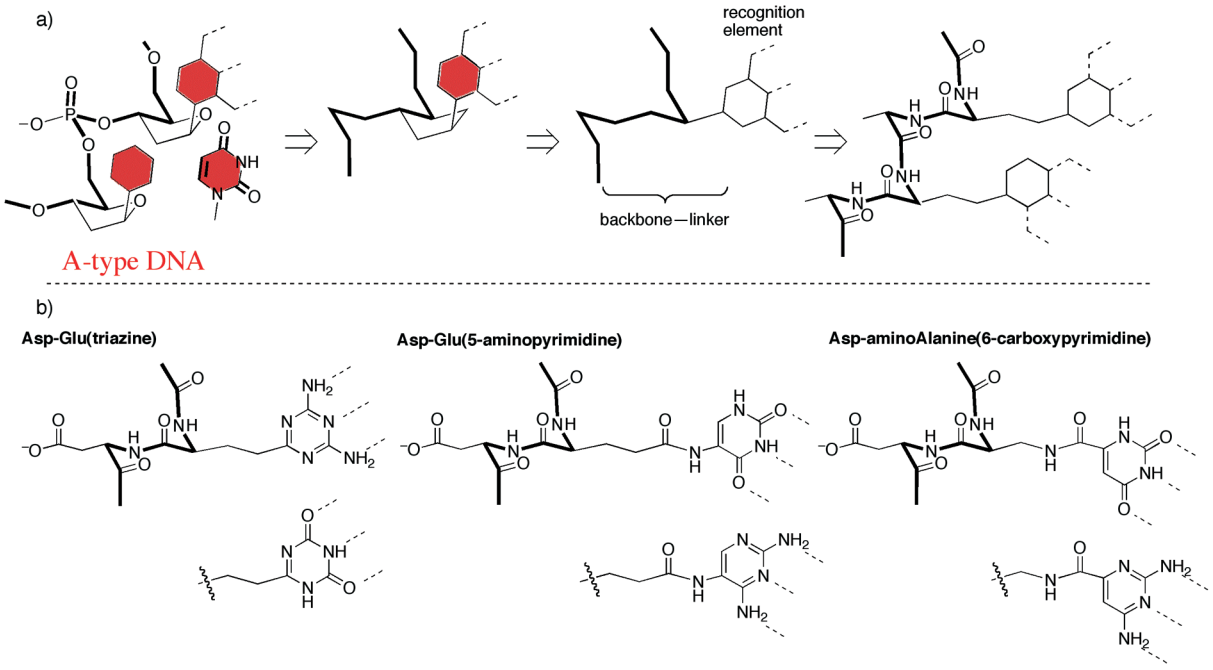

Fig. 4. General outline of a qualitative conformational analysis of an alternative oligomeric structure derived from A-type DNA (RNA) phosphodiester backbone. (a) The phosphodiester backbone (highlighted in bold) serves as a framework for the acyclic backbone scaffold. Shown on the right is one of the results of such an analysis: an oligopeptide backbone with linkers with varying length tagging the recognition element (six-member ring) to the oligomeric backbone. (b) The qualitative conformational analysis of the oligopeptide backbone tagged with the three different families of 2,4-disubsituted alternative heterocycles (with different backbone–base-pairing axis) indicate that they would be able to adopt a (idealized) conformation conducive for their functioning as informational base-pairing systems.

complementary RNA/DNA sequences containing uracil and thymine. In sharp contrast, the 2,4-dioxo-triazine tagged oligopeptide sequences showed no base-pairing with corresponding complementary RNA/DNA sequences containing adenine, and weak base-pairing with 2,6-diaminopurine containing DNA sequences. This pattern of behavior was found to be widespread among the various oligopeptide backbone constructed from amino acids such as glycine, iminodiacetic acid, aspartic acid and glutamic acid. Also, the intra-system base-pairing between 2,4-diamino-triazine and 2,4-dioxo-triazine tagged oligopeptides was found to be extremely weak.

Based on the results from the triazine family of heterocycles, we chose to study the base-pairing properties of 5-aminopyrimidines tagged to aspartyl-glutamate (Asp-Glu) oligopeptide backbone, since this would allow for a direct comparison between the two family of heterocycles tagged to the same oligopeptide backbone. We began our study with homobasic sequences of Asp-Glu oligopeptide tagged with 2,4-dioxo-5-aminopyrimidine (5-aminouracil), which showed strong base-pairing with RNA/DNA sequences containing adenine. When we examined the corresponding complementary heterocycle 2,4-diamino-5-aminopyrimidine tagged Asp-Glu oligopeptides, we were surprised to find weak base-pairing of these sequences with complementary RNA/DNA sequences containing uracil/thymine. Furthermore, weak intra-system base-pairing behavior between 2,4-dioxo-5-aminopyrimidine and 2,4-diamino-5-aminopyrimidine tagged oligopeptides was observed.

Subsequently, the intra-system base-pairing between Asp-Glu tagged triazines and 5-amino-pyrimidines was shown to follow the traits exhibited by the two stronger base-pairing partners from each family: the base-pairing strength between the 2,4-diamino-triazine and 2,4-dioxo-5-aminopyrimidine containing sequences were found to be stronger; the inverse combination, i.e. the 2,4-dioxo-triazine and 2,4-diamino-5-aminopyrimidine containing sequences was found to display weak base-pairing.

This contrasting base-pairing behavior of the different heterocycles within the two different families, while puzzling at first, became instructive when viewed in context of their physico-chemical properties and when contrasted with the corresponding properties of canonical nucleobases. The attribute that seems to best correlate with the dichotomous base-pairing behavior of these alternative heterocycles is their pKa values. For *e.g.*, 2,4-diaminotriazine (pKa ≈ 4.5), pairs strongly with thymine/uracil (pKa ≈ 9.5), with the difference in the pKa between the base-pairing partners (ΔpKa) being about 5 units; while 2,4-dioxotriazine which has a pKa ≈ 7.2, pairs weakly with adenine/2,6-diaminopurine (pKa ≈ 4.2/5), the ΔpKa being 2-3 units. Analogously, in the 5-amino-pyrimidine series, 2,4-dioxo-5-aminopyrimidine (pKa = 8.9) is the better base-pairing partner with adenine (pKa = 4.2), correlating with ΔpKa

≈ 5, while 2,4-diamino-5-aminopyrimidine (pKa ≈ 6.0) is the weaker base-pairing partner with thymine/uracil (pKa ≈ 9.5), corresponding to a ΔpKa ≈ 3.5. This (limited set of) data points to a striking ΔpKa–base-pairing strength correlation: *smaller the ΔpKa of the base-pairing partners, weaker the base-pairing strength in aqueous medium at near neutral pH*. The pKa of natural canonical nucleobases and the ΔpKa difference of about 6 among the base-pairing partners seem to correlate with their functional optimum under physiological conditions.

From an etiological viewpoint, this result —when combined with the earlier studies with different backbones— raises the possibility that in Nature's choice of nucleic acids, the constitution of the recognition component may have played a more critical role than the structure of the backbone. Such reasoning is supported by the existence of a wide variety of alternative backbones tagged with canonical nucleobases that are functionally viable, while there seems to be a dearth of (potentially natural) heterocyclic alternatives that can fulfill the role of canonical nucleobases.

2,4-dioxo-6-carboxypyrimidine tagged oligopeptides. In this family of heterocycles (derived from orotic acid), the presence of a 6-carboxyl group on the pyrimidine nucleus in combination with an amide bond tagging chemistry necessitated an amino group on the oligopeptide backbone. Therefore, we chose an aspartyl-aminoalanine (Asp-aminoAla) peptide backbone on to which the recognition elements could be attached (Fig. 4b), and focused on the synthesis of 2,4-dioxo-6-carboxypyrimidine tagged oligopeptide. Using standard synthetic methodology we obtained the building blocks necessary for automated machine assisted peptide synthesis, and assembled the 2,4-dioxo-6-carboxypyrimidine tagged hexa-peptides, dodeca-peptides and hexadeca-peptides. While in the case of hexamer we were able to procure homogeneous material, in the case of dodecamer we could only isolate mixtures containing the 10-, 11- and 12-mer; in the case of hexadecamer we isolated a mixture of 14-, 15- and 16-mer. Their identities were established by MALDI-tof mass-spectroscopy. These mixtures of oligomers, however, are suited and useful for studying their capacity for base-pairing with natural systems.

We studied the base-pairing properties of these 2,4-dioxo-6-carboxypyrimidine tagged Asp-aminoAla oligopeptides with the corresponding complementary DNA [$d(A)_{12}$, $d(D)_{12}$, $d(A)_{16}$, poly-d(A)] and RNA sequences [$r(A)_{12}$, $r(A)_{16}$, poly-r(A)] using established UV- and CD-spectroscopic methods at six different wavelengths. All of the temperature dependent UV- and CD-spectroscopic data from these studies, point to a very weak base-pairing interaction between 2,4-dioxo-6-carboxypyrimidine tagged oligopeptides and DNA/RNA at the "12-mer" and "16-mer" level (figure 5). This is also indicated by the almost negligible effect of variations in the salt concentration, and length of pairing strand, on base-pairing strength (as judged by temperature dependent UV- and CD-absorption curves). That there is indeed base-pairing interaction between 2,4-dioxo-6-carboxypyrimidine tagged oligopeptides with adenine containing oligonucleotides was shown, beyond a doubt, by a mixing curve experiment ('Job-plot', figure 6) which, not surprisingly, shows the formation of a triplex with a 2:1 ratio of 2,4-dioxo-6-carboxypyrimidine dodecapeptide with poly-r(A).

In light of our previous studies [18] where we had discovered the existence of a correlation between ΔpKa of the base-pairing partners and their base-pairing propensity, we decided to check if the observed weak pairing behavior of 2,4-dioxo-6-carboxypyrimidine is consistent with such a phenomenon. A literature survey [19] revealed that the pKa of orotamide or its close analogs were not available; therefore, we measured the pKa of the 2,4-dioxo-6-carboxypyrimidine (orotamide) unit in

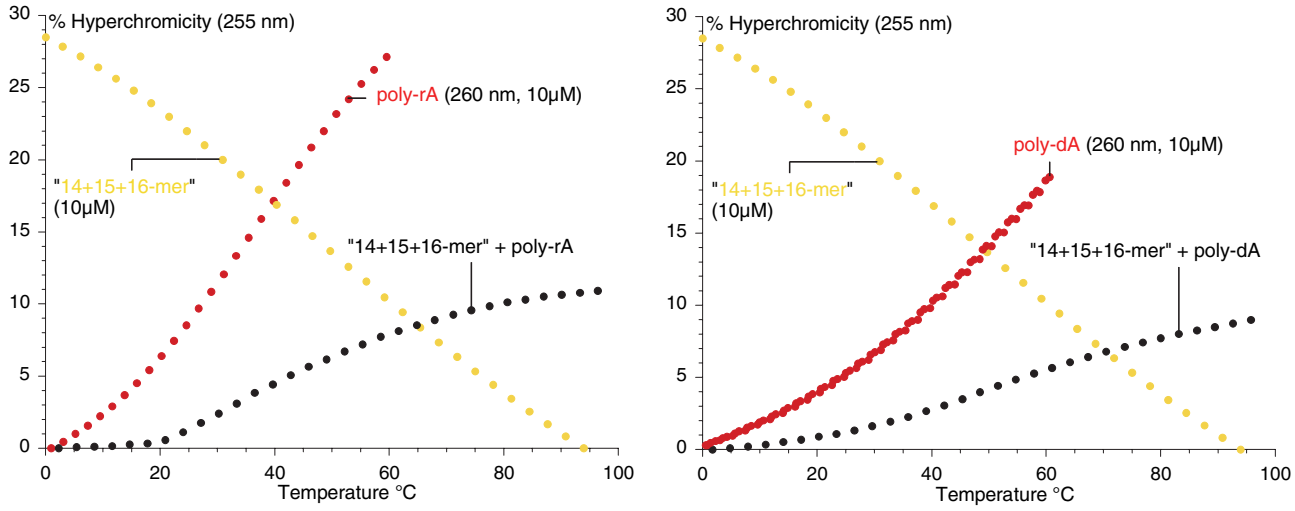

Fig. 5. UV-spectroscopic melting curves documenting weak (or no) interaction of 2,4-dioxo-6-carboxypyrimidine tagged oligopeptide ('16-mer', in yellow color) with RNA (poly-rA, in red color) and DNA (poly-dA, in red color). Conditions: ca. 5+5 μM oligomer in 1.0 M NaCl,10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7.

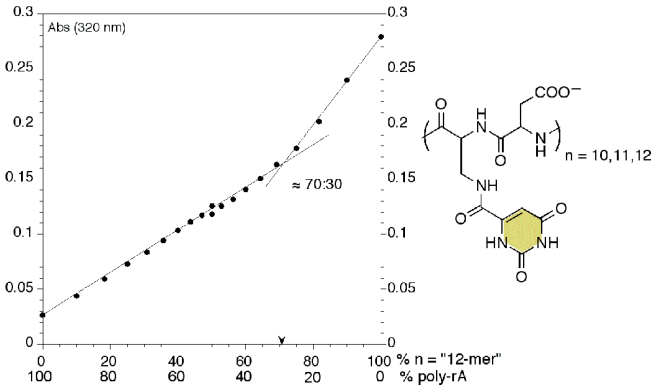

Fig. 6. UV-spectroscopic mixing curves ('Job-plot') documenting interaction of 2,4-dioxo-6-carboxypyrimidine tagged oligopeptide ('12-mer') with RNA (poly-rA). Absorbance values at 320 nm from UV curves, ca. 5+5 μM oligomers in 1M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7 at 0°C.

a Asp-aminoAla dipeptide and found it to be 6.6, a value that is lower when compared to orotic acid itself (pKa_2 = 9.45) [19], but in line with that of the methyl ester of orotic acid (pKa = 7.93 [19b]). From the standpoint of the ΔpK_a–base-pairing strength correlation criterion, the results from the above study is in accordance with the previously [18] observed correlation, '*smaller the difference in pKa between base-pairing partners, weaker the base-pairing strength*'. The difference in pKa between adenine (pK_a= 4.2) and orotamide derivative (pK_a = 6.6) is small ($\Delta pK_a \approx 2$) when compared to that of uracil/thymine (pK_a = 9.5) with adenine ($\Delta pK_a \approx 5$). Thus, when juxtaposed with canonical nucleobases and with alternative nucleobases in the pKa–base-pairing strength correlation landscape, orotamide seems to fit in (Fig. 7).

That this pKa value of 6.6 (of the 2,4-dioxo-6-carboxypyrimidine unit in a didpeptide) corresponds to the deprotonation of the N(1)-H proton (and not the N(3)-H proton) was shown by comparison of the UV-behavior of N(1)-methyl and N(3)-methyl orotic acid derivatives under deprotonating conditions [19a]. It was observed that only for the N(3)-methyl derivative (which can deprotonate only at the N(1)-H position) there was a shift in the λ_{max} from 280 to 300 nm with increasing pH of the medium [19a], a behavior that is paralleled by the 2,4-dioxo-6-carboxypyrimidine derivative unit in didpeptide (Fig. 8). Such a bathochromic shift in the UV-spectrum with increasing pH of the medium, was also observed when the temperature dependent UV-spectrum of the 2,4-dioxo-6-carboxypyrimidine tagged hexadecapeptide alone was recorded (Fig. 8). This was indicative of a temperature dependent deprotonation that was taking place on the 2,4-dioxo-6-carboxypyrimidine units within the oligomer; moreover, this same response was also observed for the 2,4-dioxo-6-carboxypyrimidine tagged oligomer in the presence of its pairing partner poly-d(A).

All these facts are indicative of the 2,4-dioxo-6-carboxypyrimidine unit undergoing deprotonation under the base-pairing measurement conditions and existing in its N(1)-anionic form; this behavior *seems* to be contributing to its weaker base-pairing strength [20]. This conclusion is supported by observations that 2,4-dioxo-5-aminopyrimidine tagged Asp-Glu oligopeptides, which have the same overall backbone–base-pairing axis (Fig. 4b), show strong base-pairing with complementary DNA and RNA sequences. In the context of our search for (potentially natural) structural alternatives of RNA, the weak base-pairing properties of 2,4-dioxo-6-carboxypyrimidine tagged oligomers discount the possibility that they could have been members of the landscape of potentially primordial *informational* oligomers.

Discussion

Based on the results from the study of base-pairing properties of these potentially natural alternative recognition elements, and pondering about the selection of canonical nucleobases over other potentially natural alternatives, one is tempted to ask, 'why nature chose these canonical nucleobases and not others'. We can, perhaps, attempt to answer it by considering the properties of the canonical bases, and by comparing them with those of the primordially relevant natural alternatives. For example, when contemplating about 'why nature chose phosphate', Westheimer [21] has pointed out the '*importance of being ionized*', emphasizing that under physiological conditions (aqueous medium & near neutral pH) the phosphodiesters are ionized, rendering them kinetically stable and enabling their retention within a bilayer membrane, thus making them suitable for their primary function as a linker in an informational polymer. Considering the pKa values of the canonical nucleobases, and by comparing them with those of the potentially natural alternative heterocycles, we are led to conclude that the exact opposite is true for the canonical nucleobases, i.e. '*the importance of being* ***not*** *ionized'* under physiological conditions.

That the pKa of canonical nucleobases (< 4 and > 9) allow them to be un-ionized at neutral pH has been highlighted in the context of emphasizing the importance of hydrophobic & stacking interactions [22] —leading to reinforcement of hydrogen bond between the appropriate base-pairing partners. The pyrimidine heterocycles (in our study) that have pKa values closer to pH of the aqueous medium are ionized (by protonation or deprotonation), and therefore, are less hydrophobic and tend to unstack. As a result, they are more prone to interacting with water molecules than with their corresponding base-pairing counterpart —leading to weak or no base-pairing interaction. Therefore, in addition to the pKa difference between base-pairing partners, the difference between pKa of the heterocycle and pH of the medium is also a factor in influencing the base-pairing capability of a given recognition element.

The results from the above study of potentially natural alternative heterocycles highlight the constitutional uniqueness of the canonical nucleobases and (once again) call attention to the relationship between their physicochemical properties and

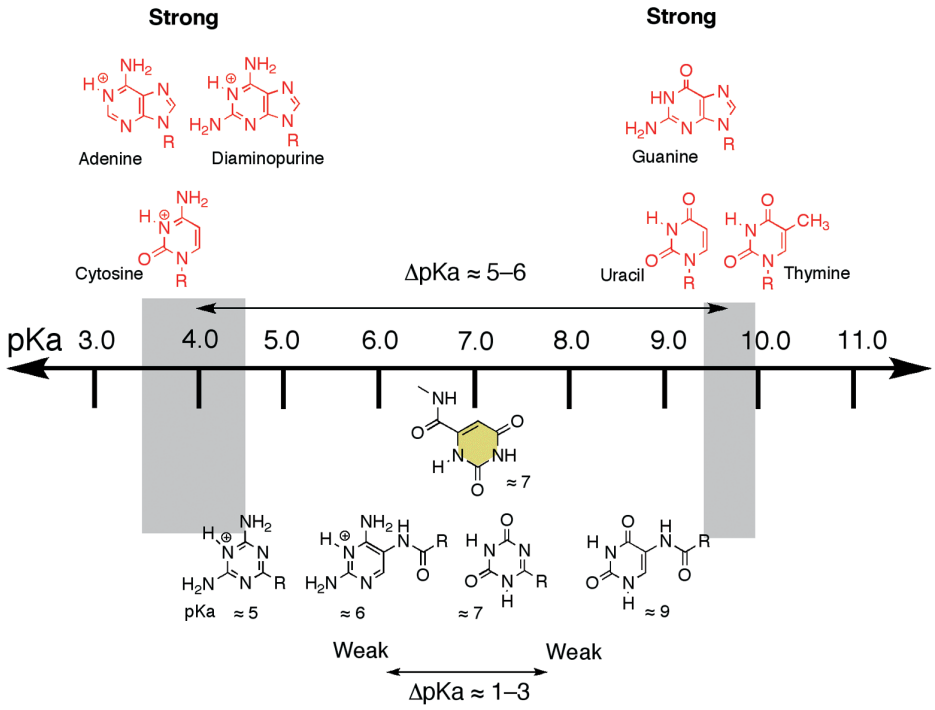

Fig. 7. pKa–Base-pairing strength correlation landscape (in near neutral aq. medium) displaying the unique positions of the canonical nucleobases with respect to the optimum difference in pKa between the base-pairing partners.

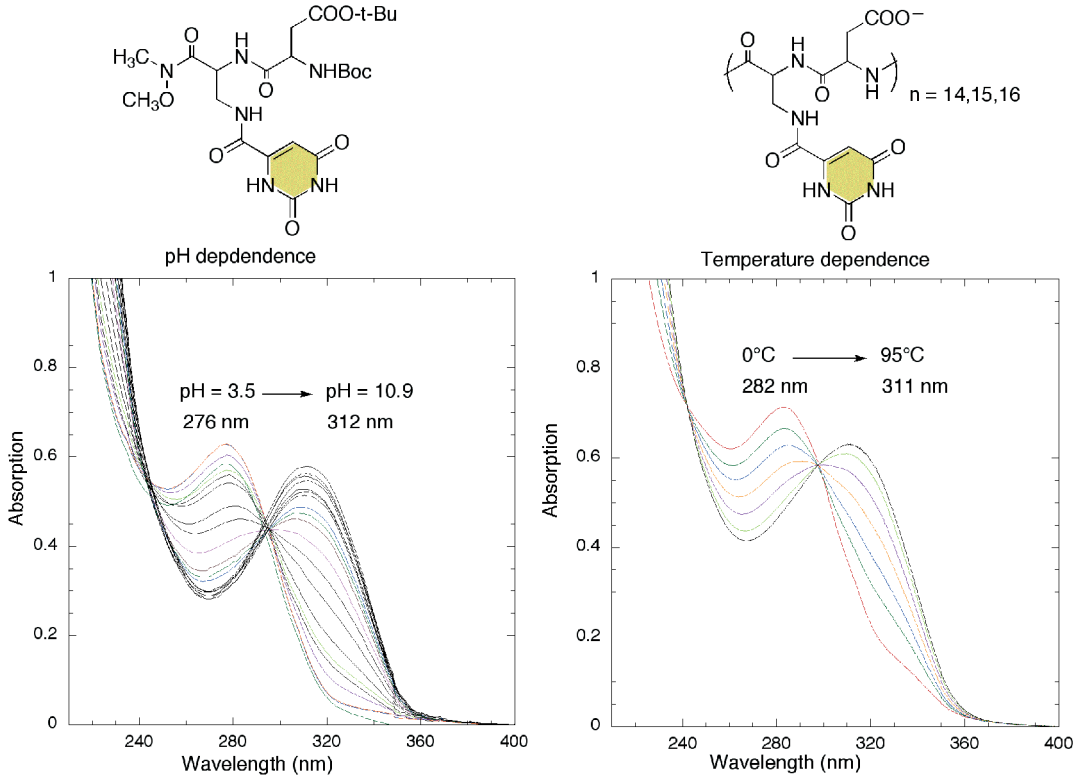

Fig. 8: UV-spectroscopic curves documenting a similar type of behavior (ascribed to deprotonation of the N(1)-H proton on the orotamide unit) that is present at the level of the monomer due to change in pH (left), and also in the oligomer due to change in temperature (right).

their functional fitness with respect to their role in informational base-pairing systems (in aqueous medium at near neutral conditions).

Acknowledgements

I am extremely grateful and indebted to Professor A. Eschenmoser for providing me the opportunity, and the privilege, to be a collaborator and a partner in the scientific endeavors in his laboratory at The Scripps Research Institute; it is an honor that I cherish. The central work presented in this lecture, the alternative heterocycle tagged oligopeptides, is the accomplishment of our talented co-workers Dr. M. Guerrero, Dr. G.K. Mittapalli, Dr. O. Munoz, Dr. Y. Osornio, Dr. K.R. Reddy, Dr. F. De Riccardis, Dr. H. Xiong, and Dr. X. Zhang. This work has been supported by The Skaggs Research Foundation at The Scripps Research Institute and by a grant (NNX07AK18G) from NASA-Astrobiology: exobiology & evolutionary biology.

References

1. A. Eschenmoser, *Science* **1999**, *284*, 2118-2124.
2. A. Eschenmoser, M. Dobler, *Helv. Chim. Acta* **1992**, *75*, 218-259.
3. A. Eschenmoser, *Origins Life Evol. Biosphere*, **1994**, *24*, 389-423.
4. (a) M. Böhringer, H.-J. Roth, J. Hunziker, M. Göbel, R. Krishnan, A. Giger, B. Schweizer, J. Schreiber, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1992**, *75*, 1416-1477. (b) J. Hunziker, H.-J. Roth, M. Böhringer, A. Giger, U. Diederichsen, M. Göbel, R. Krishnan, B. Jaun, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 259-352. (c) G. Otting, M. Billeter, K. Wüthrich, H.-J. Roth, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2701-2756. (d) K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375-474.
5. (a) S. Pitsch, S. Wendeborn, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2161-2183. (b) S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C. Lesueur, I. Schlönvogt, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1995**, *78*, 1621-1635. (c) M. Beier, F. Reck, T. Wagner, R. Krishnamurthy, A. Eschenmoser, *Science* **1999**, *283*, 699-703. (d) F. Reck, H. Wippo, R. Kudick, R. Krishnamurthy, A. Eschenmoser, A. *Helv. Chim. Acta*. **2001**, *84*, 1778-1804. (e) T. Wagner, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, A. *Helv. Chim. Acta*. **2002**, *85*, 399-416. (f) M. O. Ebert, A. Luther H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *Helv. Chim .Acta*. **2002**, *85*, 4055-4073. (g) O. Jungmann, M. Beier, A. Luther, H. K. Huynh, M. O. Ebert, B. Jaun, R. Krishnamurthy, A. Eschenmoser. *Helv. Chim. Acta*. **2003**, *86*, 1259-1308. (h) S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculca, N. Windhab, R. Micura, M. Stanek, B. Jaun, A. Eschenmoser. *Helv. Chim. Acta*. **2003**, *86*, 4270-4363.
6. (a) K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser. *Science* **2000**, *290*, 1347-1351. (b) X. Wu, G. Delgado, R. Krishnamurthy, A. Eschenmoser, *Org. Lett*. **2002**, *4*, 1283-1286. (c) C.J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *J. Am. Chem. Soc.* **2002**, *124*, 13716-13721. (d) K.-U. Schöning, P. Scholz, X. Wu, S. Guntha, G. Delgado, R. Krishnamurthy, A. Eschenmoser. *Helv. Chim. Acta*. **2002**, *85*, 4111-4153.
7. P. E. Nielsen, *Acc. Chem. Res*. **1999**, *32*, 624-630.
8. (a) P. Herdwijn, *Liebigs Ann. Chem*. **1996**, *9*, 1337-1348. (b) B. Hyrup, P. Nielsen, *Bioorg. Chem. Med. Chem. Lett.* **1996**, *4*, 5-13. (c) J. Micklefield, *Current Med. Chem*. **2001**, *8*, 1157-1179.
9. A. Eschenmoser, *Origins Life Evol. Biosphere*, **2004**, *34*, 277-306.
10. A. Eschenmoser, *Chimia*, **2005**, *59*, 836-850.
11. (a) R. A. Sanchez, L. E. Orgel, *J Mol. Biol*. **1970**, *47*, 531-543. (b) W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol*. **1972**, *1*, 249-257. (c) G. Zubay, T. Mui, *Origins Life Evol. Biosphere* **2001**, *31*, 87-102.
12. (a) Z. Wang, H. K. Huynh, B. Han, R. Krishnamurthy, A. Eschenmoser. *Org. Lett*. **2003**, *5*, 2067-2070. (b) B. Han, Z. Wang, B. Jaun, R. Krishnamurthy, A. Eschenmoser. *Org. Lett*. **2003**, *5*, 2071-2074. (c) B. Han, B. Jaun, R. Krishnamurthy, A. Eschenmoser. *Org. Lett*. **2004**, *6*, 3691-3694.
13. (a) C. G. Overberger, F. W. Michelotti, P. M. Carabateas, *J. Am. Chem. Soc*., **1957**, *79*, 941-944. (b) H. Bredereck, F. Effenberger, A. Hofmann, M. Hajek, *Angew. Chem Int. ed. Eng*. (review), **1963**, *2*, 655-659. (c) J. K. Simons, M. R. Saxton, *Organic Synthesis*, Wiley, New York, 1963, Collect. Vol. IV, pp78-79. (d) H. Bredereck, F. Effenberger, A. Hoffmann, *Ber*., **1964**, *97*, 61-63. (e) E. Kuwano, E. Taniguchi, K. Maekawa, *Agric. Biol. Chem.*, **1973**, *37*, 423-430. (f) R. Deans, G. Cooke, V. M. Rotello, *J. Org. Chem*., **1997**, *62*, 836-839. (g) Z. Brozozowski, F. Saczewski, M. Gdaniec, *Eur. J. Med. Chem*., **2000**, *35*, 1053-1064. (h) A. D-Ortiz, J. Elguero, C. F-Foces, A. d-l. Hoz, A. Moreno, M. d-C. Mateo, A. S-Migallon, G. Valiente. *New. J. Chem.*, **2004**, *28*, 952-958.
14. (a) J. P. Ferris, P. C. Joshi, *Science*, **1978**, *201*, 361-362. (b) A. B. Voet, A. W. Schwartz, *Origins Life Evol. Biosphere*, **1982**, *12*, 45-49. (c) S. Miyakawa, H. J. Cleaves, S. L. Miller, *Origins Life Evol. Biosphere*, **2002**, *32*, 209-218.
15. J. P. Ferris, P.C. Joshi, E. H. Edelson, J. G. Lawless, *J. Mol. Evol*. **1978,** *11*, 293–311.
16. (a) N. K. Sharma, K. N. Ganesh, *Chem. Commun*. **2003**, 2484-2485. (b) T. H. S. Tan, D. T. Hickman, J. Morral, I. G. Beadham, J. Micklefield, *Chem. Commun*. **2004**, 516-517; (c) P. von Matt, A. De Mesmaeker, U. Pieles, W. Zürcher, K-H Altmann, *Tetrahedron Lett*. **1999**, *40*, 2899-2902. (d) T. Vilaivan, G. Lowe, *J. Am. Chem. Soc*. **2002**, *124*, 9326-9327. (e) R. A. Goodnow, S. Tam, D. L. Pruess, W. W. McComas, *Tetrahedron Lett*. **1997**, *38*, 3199-3202. (f) M. Kuwahara, M. Arimitsu, M. Shigeyasu, N. Saeki, M. Sisido, *J. Am. Chem. Soc*. **2001**, 123, 4653-4658. (g) A. Lenzi, G. Reginato, M. Taddei, E. Trifilieff, *Tetrahedron Lett.* **1995**, *36*, 1717-1718. (h) U. Diederichsen, *Angew. Chem. Int. Ed.* **1996**, *35*, 445-448. (i) A. M. Brückner, H. W. Schmitt, U. Diederichsen, *Helv. Chim. Acta*, **2002**, *85*, 3855-3866. (j) M. Fujii, K. Yoshida, J. Hidaka, T. Ohtsu, *Chem. Commun*. **1998**, 717-718. (k) Y. Huang, S. Dey, X. Zhang, F. Sönnichsen, P. Garner, *J. Am. Chem. Soc*. **2004**,*126*, 4626-4640. (l) H. Umemiya, K. Komatsu, T. Yamzaki, H. Kagechika, K. Shudo, Y. Hashimoto, *Nucl. Acid, Symp. Ser.* **1995**, *34*, 37-38.
17. 2,3-Diaminoacids have been identified in Murchison meteorite (U. J. Meierhenrich, G. M. Munoz Caro, J. H. Bredehöft, E. K. Jessberger, W. H.-P. Thiemann, *Proc. Natl. Acad. Sci*. **2004**, *101*, 9182-9186), in laboratory experiments simulating chemistry of interstellar medium (G. M. Munoz Caro, U.J. Meierhenrich, W. A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H.-P. Thiemann, A. Brack, J. M. Greenberg, *Nature*, **2002**, *416*, 403-406) and as a product of hydrolysis of NH4CN solutions (K. E. Nelson, M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci*. **2000**, *97*, 3868-3871).
18. (a) G. K. Mittapalli, K. R. Reddy, H. Xiong, O. Munoz, B. Han, F. De Riccardis, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem.*

Int. Ed. **2007**, *46*, 2470-2477. (b) G. K. Mittapalli, Y. M. Osornio, M. A. Guerrero, K. R. Reddy, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem. Int. Ed*. **2007**, *46*, 2478-2484.

19. (a) J. J. Fox, N. Yung, I. Wempen, *Biochim. et Biophys. Acta*, **1957**, *23*, 295-305. (b) J. Jonas, J. Gut, *Collect. Czech. Chem. Commun*. **1962**, *27*, 716-723. (c) M. S. Masoud, M. Abd El Zaher Mostafa, R. H. Ahmed, N.H. Abd El Moneim, *Molecules*, **2003**, *8*, 430-438.
20. It is not excluded that there may be other reasons, e.g. conformational restrictions, that are also responsible for the weak pairing of 2,4-dioxo-6-carboxypyrimidine tagged oligopeptides. We are currently checking this possibility by making the complementary 2,4-diamino-6-carboxypyrimidine tagged oligopeptide (which has a pKa of about 4.6, and is expected to show base-pairing with thymine/uracil containing oligomers).
21. F. H. Westheimer, *Science* **1987,** 235, (4793), 1173-1178.
22. (a) P.O.P. Ts'o, *Ann. N. Y. Acad. Sci*. **1969**, *153*, 785-804. (b) W. Saenger, "Principles of Nucleic Acid Structure", 1984, ed. C. R. Cantor, Springer Verlag, New York, N.Y. (c) M. Roitzsch, B. Lippert, *J. Am. Chem. Soc*. **2004**, *126*, 2421-2424.