Pure & Appl. Chem., Vol. 63, No. 10, pp. 1357–1366, 1991.
Printed in Great Britain.

The thermodynamic basis of the genetic code

H.H.Klump and D.L.Maeder

Department of Biochemistry, University of Cape Town, Rondebosch, 7700 South Africa.

Abstract: A systematic investigation of the helix-coil transition of DNA's as a function of sequence composition has lead to the conclusion that the Gibbs energy per base pair is a zero-order property. There is a correlation between the observed codon usage frequency and the codon/anti-codon interaction Gibbs energy per triplet. We will show that the bias of codon usage reflects strategies for efficiency and accuracy in gene expression. Organization beyond the codon triplets is discussed.

THE TWO KINGDOMS: THERMODYNAMICS AND GENETICS

The great abstract value of thermodynamics stems from the fact that since the change of a system from its initial state to the final state does not depend on the path taken, we do not need to know anything about the mechanism. The great practical use of thermodynamics originates from the fact that changes in state functions such as energy or entropy do not depend on a particular path taken from one state to another, and therefore we will always be able to equate the end results of proceeding from one state (1) to another by two different paths. We can thereby obtain information about reactions that we could not study directly, such as those taking place within living cells.

Specifically, thermodynamics gives a quantitative basis for predicting energy changes resulting from chemical reactions or accompanying conformational changes. This information is crucial in determining the reaction routes involved in various metabolic pathways. The full formal elegance of thermodynamics becomes evident from a careful development of the first and second laws, which is outside the scope of this discussion. However, there is a practical desire for thermodynamic data because they define the nature of the inter- and intra-molecular forces that stabilize a particular conformation of a biopolymer (2-4) . This interest intensified when it was realized that such data can serve as the basis for predicting sequence-specific secondary structural domains along linear polymer chains, such as DNA. DNA is the main creature in the other Kingdom, genetics.

The formal elegance of thermodynamics is completely matched by that of the concept of the genetic code (5). After Avery's demonstration that DNA is the genetic material, Chargaff's analyses showing that the relative ratios of the four bases in genomic DNA were far from random, and Watson and Crick's elegant demonstration of the

structure of the double helical conformation, it was only a matter of time before it was proposed that groups of three nucleotides (triplets) could serve the purpose of coding for the twenty amino acids found in almost all protein sequences (6).

The code must be degenerate because the 64 permutations provide much more code words than needed for a unique correlation of a triplet and a specific amino acid. The experimental breakthrough in identifying which triplet determines which specific amino acid in a sequence came when Matthei and Nirenberg demonstrated that the synthetic messenger RNA poly-uridylic acid codes for the monotonous polypeptide poly-phenylalanine (7).

Use of increasingly more complex polynucleotides as synthesized in Khorana's laboratory finally established the complete set of codons. It turns out that 61 out of 64 possible triplets correspond to amino acids, most of them being coded by more than one, and the residual three codons being stop signals. Consequently the smallest unit of the DNA sequence which is relevant in the genetic context is the triplet and not the single nucleotide. In non-coding regions of the total genome, other unit sizes may be considered (8).

The aim of our approach is to combine the two Kingdoms in such a way that we can assign a characteristic thermodynamic parameter to any triplet. The parameter chosen is the interaction Gibbs energy between each codon and its anti-codon. In the following we will demonstrate that there is indeed a correlation between interaction Gibbs energy and the preference for certain codons in certain genes (9).

EXPERIMENTAL APPROACH TO THE INTERACTION GIBBS ENERGY

The concerted thermal denaturation of a long DNA sequence, normally termed helix coil transition, is a sequence-dependent process. The H-bonds in the plane of the individual base pairs have to be broken; a process similar to the destabilization of H-bonds in melting ice. Accordingly, the temperature at which 50% of the bases are unpaired is called the melting temperature (Tm)(10). The melting of DNA can conveniently be followed by recording the change in absorbance at 260 nm due to the unstacking of consecutive bases in the single strands or of partly overlapping bases in opposite strands. Stacking interactions contribute to the overall helix stability as much as the intact H-bonds (11).

The helical structure is destabilized by electrostatic interactions between the backbone phosphate groups as well as by the conformational entropy decrease resulting from the freezing of free rotation around single bonds along the two backbones. The entropy term is expected to be sequence-independent. Since no bonds inside a particular base are involved, it is dependent only on the total number of degrees of freedom restricted upon helix formation. Using Boltzmann's formalism (12) $S = n.R.\ln(w)$ to calculate the entropy change per subunit (nucleotide) the maximal entropy change should amount to 13 e.u. because in our case we have to consider six single bonds along the DNA backbone (n=6) and three preferred conformations per bond (w=3). As will be shown later, the experimental results fit very well with the above assumption (Fig. 1c).

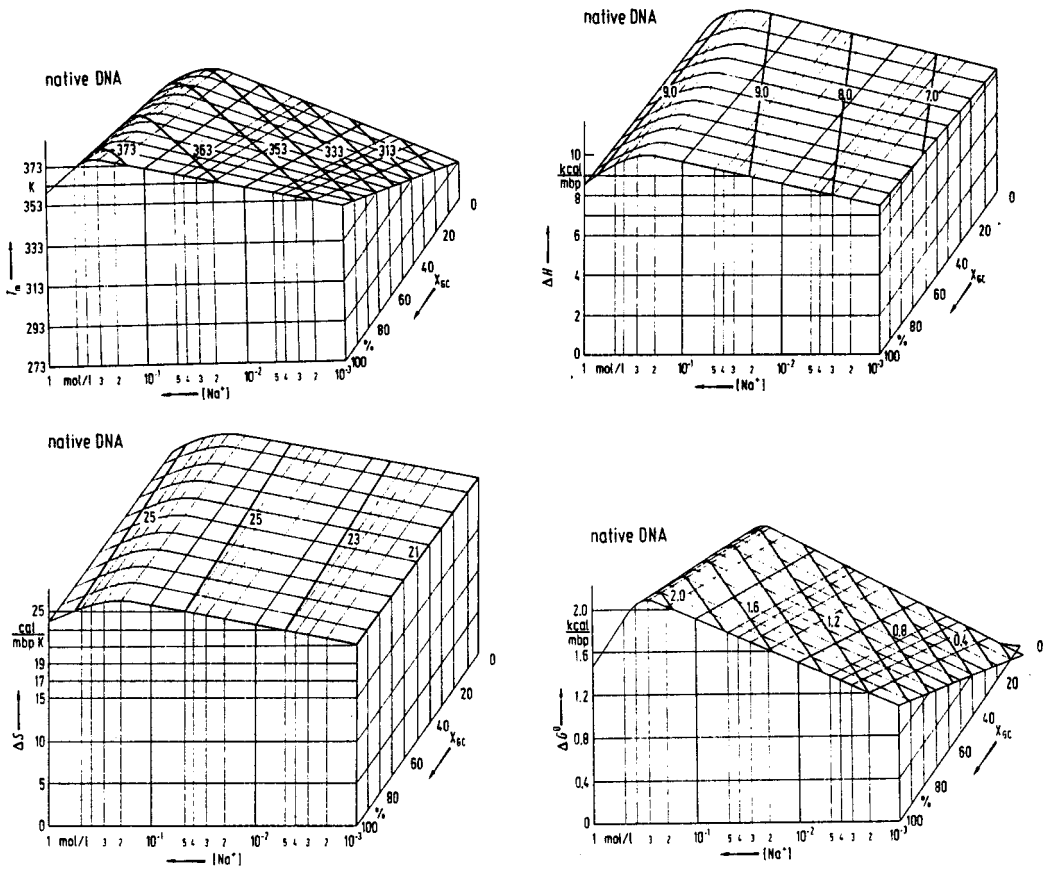

Fig. 1 a-d. Thermodynamic state functions for all DNAs.

The impact of the electrostatic interactions shows up in the ionic strength dependence of Tm. A tenfold increase of the cation concentration shifts the Tm about 16° upwards. This shift is linear between millimolar and 0.3 molar cation concentration. The sequence and solvent dependence of the four thermodynamic parameters Tm, enthalpy, entropy, and Gibbs energy are shown in Fig. 1 a-d.

Enthalpy, entropy, and Gibbs energy changes are calculated per mol of base pairs. For a given ionic strength, the four parameters change linearly with the gross GC content of the sample DNA. The properties listed are zero order properties, i.e. the stability of an individual base pair is independent of its particular nearest neighbor base pairs. An AT-pair next to a GC-pair is not more stable than an AT-pair next to another AT-pair (13).

The influence of nearest neighbors on the energetics of base pairing can be tested by using a selected body of nucleotides, namely six, to account for all possible next nearest neighbor interactions. Poly-dA poly-dT and poly(dA-dT) will account for the two possible pure AT sequences, the first representing the homo-purine/homo-pyrimidine helices, the second representing the strictly alternating purine/pyrimidine sequences. Poly-dG poly-dC and poly(dG-dC) will represent the two alternative GC environments. Finally the mixed AT/GC sequences will be adequately modelled by poly(dA-dG) poly(dC-dT), and poly(dA-dC)-poly(dG-dT) respectively. Fig. 2 shows the linear dependence of the transition enthalpy per mol base pairs on the sequence composition. Obviously the

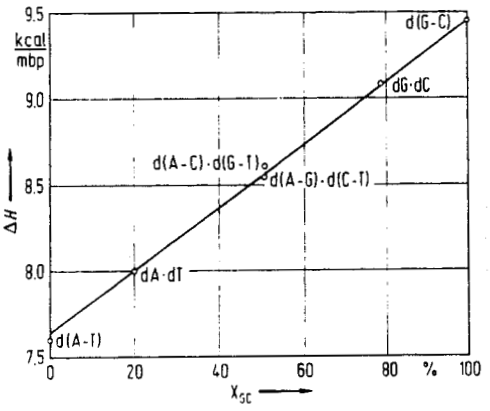

Fig. 2.
Transition enthalpy per mol base pairs vs. sequence composition

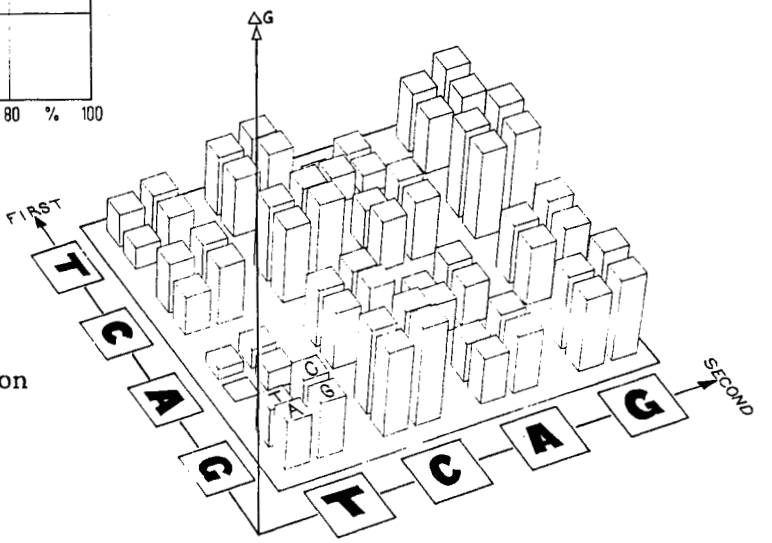

Fig. 3.
The codon/anti-codon Gibbs energy

homopurine/homopyrimidine complex of ATs is more stable than the complementary alternating sequence, but the opposite holds for the GC sequences. The 50% mix of ATs and GCs is only marginally dependent on the orientation of the purines or pyrimidines respectively. Furthermore the value obtained for the synthetic polynucleotides coincides with the value of *E.coli* DNA (per mol base pairs), which is in fact the natural 50% mix.

These results as a whole justify the summation of the individual energies (per mol base pairs) for the three bases of each codon and present the codon anti-codon interaction Gibbs energy per mol of base triplets (Fig. 3). The base of the diagram is the 4 by 4 matrix originally developed to represent the genetic code, i.e. the combination of T, C, A, or G in the first position with T, C, A, or G in the second position gives sixteen squares, each of them carrying four columns to account for either T, C, A, or G in the third position. The height of each column represents the magnitude of the interaction Gibbs energy.

The standard Gibbs energy change for the triplet ATA amounts to 0.0 kJ/mol triplet, in other words, poly(dA-dT) melts at 25°C. Alternating AT sequences are the least stable of all DNA sequences. In contrast to AT, alternating GC sequences are the most stable among the DNA sequences. Their codon/anti-codon binding Gibbs energy amounts to almost 25 kJ per mol triplet.

At first glance one can see that there are four islands of exceptionally large interaction energy, namely the triplets GCX, CGX, GGX, and CCX. They represent the code quartets for Ala, Arg, Gly and Pro, which are among the most abundant amino acids. It suggests that the best seats, so to speak, were taken first.

CORRELATION BETWEEN OBSERVED CODON FREQUENCIES AND INTERACTION GIBBS ENERGIES

The genetic code is degenerate, but alternative synonymous codons are generally not used with equal frequency. The pattern of bias is broadly similar in genes from a single species or taxonomic group. In the two best studied organisms, namely *E.coli* and *Saccharomyces cerevisiae*, genes differ largely in the degree of bias rather than its direction (14-17).

These data may provide evidence suggesting the generality of certain phenomena, e.g. the high speed of transcription and replication in unicellular organisms or the single strandedness of certain phage DNAs. As will be discussed in some detail here, closely related organisms have similar patterns of codon usage: e.g. *Salmonella typhimurium* closely resembles *E.coli*, while all mammalian species so far examined (practically mouse, rat and cow) largely resemble humans. Also included in our considerations are plants, a fungus with an extremely low GC content, and single-stranded phage DNAs (Table 1).

Various explanations have been offered in the literature of the obvious fact that codon usage is highly biased. The different concepts put forward can be classified as:

1. the "optimization of translation kinetics" concept;
2. the "need to have a uniform double helix stability" concept;
3. the "mRNA secondary structure" or "transcriptional efficiency" concept; and
4. the "most abundant tRNA codon choice" or "translational efficiency" concept. (18-21)

We favour a variation of the second concept which we will call the "codon/anticodon interaction Gibbs energy" concept. It can be demonstrated to underlie some of the other concepts as well. Looking at interaction Gibbs energy can serve to explain various strategies. In the following we will discuss three observed strategies in greater detail.

Eukaryotes are slow and stringent

The first strategy is observed for eukaryotes, and a clear picture emerges from viewing codon preferences listed for human genes. Humans tend to follow Oscar Wilde's definition of a simple taste; they only go for the best. Choosing codons with highest interaction Gibbs energies is equivalent to choosing the most accurate pairing, especially with respect to the third (wobble) position. The most accurate pairing is equivalent to the longest residence time for the codon/anticodon reading. This inevitably leads to the slowest transcription rate (22). Mammals/humans can cope with it. Bacterial gene sequences engineered to contain these preferred eukaryotic codons are transcribed more slowly than the same gene sequence composed from preferred bacterial codons. There are variations of this strategy, but the take-home message seems to be: eukaryotes go for precision. Their prime aim is to keep the large genome as intact as possible.

But there are exceptions to the rule. The principle seems to be violated in the eukaryotic codon choice when there is a C followed by a G either within the same triplet, or, though less pronounced, between triplets. It was observed that most higher eukaryotes (but

not insects) show a selective deficiency in the dinucleotide pair 5'C-G 3'. The proposed explanation for this deficiency is that these C residues are methylated and as a consequence hyper-mutation of the methyl-C (23-24) leads to the elimination of C-G dinucleotides. Inspecting human codon frequency tables shows that the two Arg codons AGA and AGG, for example, are preferred over the "best" Arg codons CGC and CGG, because of the CG neighborhood in the latter. The preferred Ser codon is UCC instead of UCG,and this avoidance of intracodon CGs is also obvious from the selection rules visible in the following three quartet codons, such as the preference of the codon ACC instead of ACG for Thr, CCC over CCG for Pro, and GCC over GCG for Ala.

Table 1. Preferred codon usage in a. *E.coli*, b. *B.subtilis*, c. *T.ferrooxidans*, d. ssDNA, e. yeast, f. maize, g. human

	Source	a	b	c	d	e	f	g
ARG	CGA	-	-	-	-	-	-	-
	CGC	-	-	CGC	-	-	-	-
	CGG	-	-	-	-	-	-	-
	CGU	CGU	CGU	-	CGU	-	-	-
	AGA	-	-	-	-	AGA	-	-
	AGG	-	-	-	-	-	AGG	AGG
LEU	CUA	-	-	-	-	-	-	-
	CUC	-	-	-	-	-	-	-
	CUG	CUG	-	CUG	-	-	CUG	CUG
	CUU	-	CUU	-	CUU	-	-	-
	UUA	-	-	-	-	-	-	-
	UUG	-	-	-	-	UUG	-	-
SER	UCA	-	-	-	-	-	-	-
	UCC	-	-	UCC	-	-	UCC	UCC
	UCG	-	-	-	-	-	-	-
	UCU	UCU	UCU	-	UCU	UCU	-	-
	AGC	-	-	-	-	-	-	-
	AGU	-	-	-	-	-	-	-
THR	ACA	-	-	-	-	-	-	-
	ACC	ACC	-	ACC	-	ACC	ACC	ACC
	ACG	-	-	-	-	-	-	-
	ACU	-	ACU	-	ACU	-	-	-
PRO	CCA	-	-	-	-	CCA	-	-
	CCC	-	-	CCC	-	-	CCC	CCC
	CCG	CCG	-	-	-	-	-	-
	CCU	-	CCU	-	CCU	-	-	-
ALA	GCA	-	-	-	-	-	-	-
	GCC	-	-	GCC	-	-	GCC	GCC
	GCG	-	-	-	-	-	-	-
	GCU	GCU	GCU	-	GCU	GCU	-	-
GLY	GGA	-	GGA	-	-	-	-	-
	GGC	-	-	GGC	-	-	GGC	GGC
	GGG	-	-	-	-	-	-	-
	GGU	GGU	-	-	GGU	GGU	-	-
VAL	GUA	-	-	-	-	-	-	-
	GUC	-	-	-	-	-	-	-
	GUG	-	-	GUG	-	-	GUG	GUG
	GUU	GUU	GUU	-	GUU	GUU	-	-
LYS	AAA	AAA	AAA	-	AAA	-	-	-
	AAG	-	-	AAG	-	AAG	AAG	AAG
ASN	AAC	AAC	AAC	AAC	-	AAC	AAC	AAC
	AAU	-	-	-	AAU	-	-	-
GLN	CAA	-	CAA	-	CAA	CAA	-	-
	CAG	CAG	-	CAG	-	-	CAG	CAG
HIS	CAC	-	-	CAC	-	CAC	CAC	CAC
	CAU	CAU	CAU	-	CAU	-	-	-
GLU	GAA	GAA	GAA	GAA	GAA	GAA	-	-
	GAG	-	-	-	-	-	GAG	GAG
ASP	GAC	GAC	GAC	GAC	-	GAC	GAC	GAC
	GAU	-	-	-	GAU	-	-	-
TYR	UAC	-	UAC	UAC	-	UAC	UAC	UAC
	UAU	UAU	-	-	UAU	-	-	-
CYS	UGC	UGC	UGC	UGC	-	-	UGC	UGC
	UGU	-	-	-	UGU	UGU	-	-
PHE	UUC	-	UUC	UUC	-	UUC	UUC	UUC
	UUU	UUU	-	-	UUU	-	-	-
ILE	AUA	-	-	-	-	-	-	-
	AUC	AUC	AUC	AUC	-	AUC	AUC	AUC
	AUU	-	-	-	AUU	-	-	-
MET	AUG	AUG	AUG	AUG	AUG	AUG	AUG	AUG
TRP	UGG	UGG	UGG	UGG	UGG	UGG	UGG	UGG
%GC		48/50	41/42	62/60	31/-	44/42	63/42	63/42

We can also include the intercodon CG dinucleotides in the discussion. Inspecting the published codon preference tables of eukaryotes shows that the frequency of codons NNC is 0.35 and GNN is 0.31. The expected intercodon frequency of CG is thus 0.11, whereas the observed frequency is 0.055. For example the codon utilization of histidine is 58% CAC and 42% CAU. This distribution reverses if the following codon starts with G. The frequency of CAC drops to 37% and the frequency of CAU rises to 63% (25).

There is no obvious tendency for the selection of the third base of a codon to offset the general GC content of the first two bases of each codon to produce a balanced GC content. The percentage GC of the total human genome is about 42% GC, while the mean value calculated from the list of preferred codons amounts to 63%. This tendency is even visible for an extreme case (26), the genome of the fungus *Dictyostelium discoideum*, which has a Tm derived GC content of only 25% but a codon preference based GC content of 35%.

Prokaryotes are fast and flexible

The alternative strategy is obviously to choose codons with intermediate stability. Precision may be sacrificed but what is gained is speed. Selecting codons with intermediate interaction energies leads to a gene with intermediate and uniform stability around an individual characteristic stability value. The particular codon preference is biased by the gross GC content of the gene sequence chosen for the investigation. The third base of each codon in the coding sequences of bacteria plays a decisive role in producing the homogeneous stability distribution along the entire gene sequence(27): the redundancy at the wobble third base counterbalances the G+C content variation at the first and second bases of preferred codons. This can be readily demonstrated by calculating the mean GC content of the eighteen preferred codons (not counting the codons for Met and Trp). The mean value calculated from codon preferences for *E.coli* amounts to 48.2% GC compared to 50% for the total *E.coli* genome. The mean value for *B.subtilis* amounts to 41.2% GC calculated from the codon preference list (Table 1), which is almost exactly equal to the 42% GC calculated from Tm for the total genomic DNA. The results for *T.ferrooxidans*, which has a Tm derived GC content of 62%, yield 59% GC content computed from the list of preferred codons for this bacterial genome. The preferred codons are basically always codons of intermediate stability and the frequency of G or C in the third position of each of the preferred codons serves, in addition, to counterbalance the average local GC frequency at the first and second sites in a gene. This strategy is adopted by prokaryotes and mainly serves the purpose of trading accuracy in gene expression, which is so important for the large genomes of eukaryotes, with high efficiency and speed, whenever the nutrition supply becomes favorable (28). There is no avoidance of CG neighborhoods. Neither intercodon nor intracodon CGs are selected against. Whatever the reason for the avoidance of this strategy in higher eukaryotes may be (possibly some regulatory role of the methylated/ nonmethylated C adjacent to the G), this selective pressure is absent in prokaryotes.

In unicellular organisms, the highly expressed genes use a smaller subset of codons as compared to the genes which are more weakly expressed. A survey of 165 *E.coli* genes reveals a positive correlation between high expression and increased codon bias. A

similar trend is described for the codon selection in yeast. Codon usage in the highly expressed genes correlates with the abundance of iso-accepting tRNAs in both *E.coli* and yeast. The good fit of the preferred codon usage with the iso-acceptor tRNA abundance explains the observed high translation level and the high steady-state level of these proteins. The resulting correlation hints at convergent evolution of the tRNA genes and the species-specific codon choice. This codon dialect as it was called by Ikemura(29), however, is not the consequence of the abundance of a given set of abundant tRNAs. Any correlation between the codon usage profile and the secondary structure of the transcribed mRNA is only a chance event. Currently we are unable to put forward any rule correlating mRNA secondary structure, let alone tertiary structure or translation efficiency.

Single-strand DNA viruses just like to stay single

The third type of straightforward strategy can be derived from the list of preferred codons observed for the single-stranded viruses. They obviously select for the least stable codon/ anticodon pairs, unrelated to the GC content of their hosts, to secure their single-strandedness in the first place. If one expands the analysis to plants, the observation is that the nuclear genome follows the first strategy, while the chloroplast follows the second strategy. More specifically, one finds that monocots share the most commonly used codon for only one of eighteen amino acids, while dicots share the most commonly used codons of chloroplasts for only four of eighteen amino acids.(30) In general, the chloroplast codon profile resembles that of unicellular organisms, with a strong bias towards the use of A+T in the wobble position.

CORRELATION BETWEEN THE PHYSICAL PROPERTIES OF A REGION OF DNA AND ITS GENETIC FUNCTION

Virtually all important biological processes on DNA, such as replication, transcription, and recombination, require partial or total disruption of the double helix. The helix-coil transition was studied intensively using a variety of spectroscopic methods, the change in UV absorbance at 260 nm as a function of temperature (12-13), CD spectroscopy, Raman/IR spectroscopy or NMR-techniques.

The latest technique in the field of high resolution thermal denaturation studies is microcalorimetry (4). Improved instrumentation and data interpretation justify the optimism for our understanding of segmental melting of DNA (31). Melting curves of total genomic DNA of certain mammals such as Bovidae (domestic cattle) or African antelopes (fig. 4), exhibit an asymmetric main melting peak with up to four distinctive superimposed sharp peaks (32-33), representing relatively short sequences, present in high copy numbers (satellite DNAs). From the position on the temperature scale and from the half-width of the extra peaks, one can determine the composition (GC%) and the mean length of these inter-dispersed sequences. This opens a new way to the determination of satellites in genomic DNAs and to taxonomic classification based on the composition or frequency of characteristic satellite sequences.

Recent thermodynamic investigations on the melting of plasmidic DNA reveal a highly segmented composition of the plasmids, reflected in a series of local sequence- specific helix-coil transitions (fig. 5). This is a remarkable characteristic of the plasmid DNA

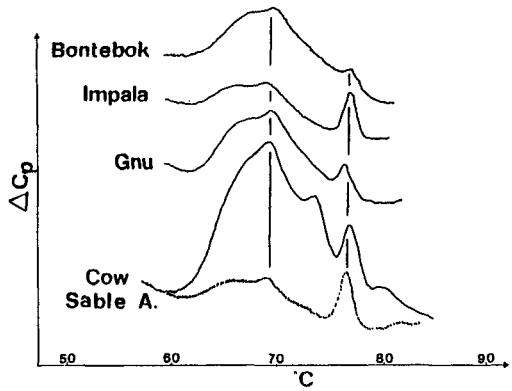

Fig. 4.
Calorimetric transition curves for African antelopes

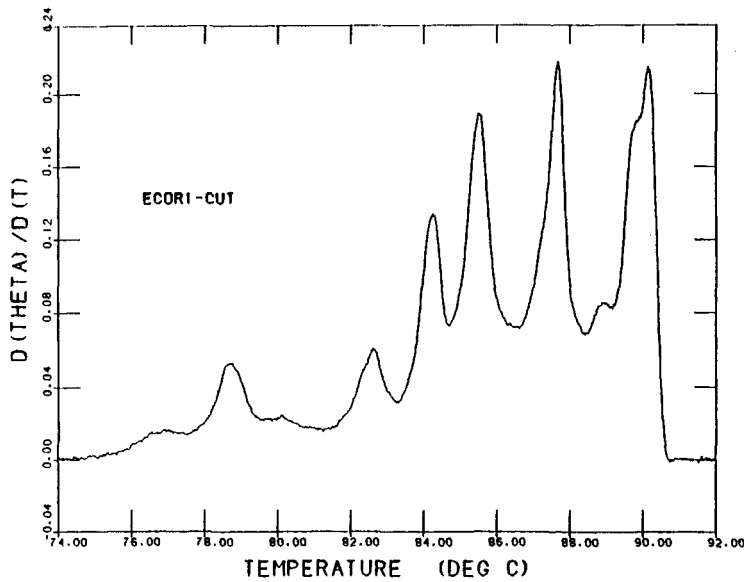

Fig. 5.
High resolution melting curve of a linearized plasmid (data from ref. 32)

double helix; it consists of virtual blocks with different degrees of physical stability, whose length is in the range of a hundred to several hundred base-pairs (34). Wada et al (27) have termed these blocks a cooperative melting region (CMR). Deviating from the mammalian genome, where the CMRs represent only a fraction of the total, it is possible to compose the entire melting curve from superimposed sub-transitions. The distribution pattern of these sub-transitions represents the unique fingerprint of a given plasmid. By comparing a series of related plasmids it is possible to detect the parental plasmid, the engineered sequence, and its impact on its flanking sequences.

The object of these studies is to correlate the physical map with the genetic map. Some preliminary results point in a very interesting direction. It appears that sequences which code for a certain proteins seem to consist of codons with intermediate and uniform stability, while sequences which according to their composition are identified as signal sequences are located in a less uniform and thermally less uniform or stable environment. Small variations in the environmental conditions will perturb these sequences much more than they would affect coding sequences. The sequence responds to the change in the environment by amplifying the perturbation in a localized sequence, exactly what one would expect from a signal transducer. Coding sequences, however, do not respond to changes as long as the perturbation intensity stays below a certain threshold value.

Further investigations are required to consolidate these emerging lines of evidence. The final goal will be to predict the physical and the correlated genetic features of a DNA sequence from its sequence. Now we can read the book of nature, letter by letter, then we will spell sentences and paragraphs, and become literate in the genetic text and the true meaning of the words.

Acknowledgements We would like to thank P.Becker for writing the program to generate Fig. 3. The constant and generous support of the Foundation for Research Development (S.A.) is gratefully acknowledged.

REFERENCES

1. Lesk,A. Introduction to Physical Chemistry, 57-59, Prentice Hall Eaglewood Cliff, NY (1982).
2. K.Breslauer, R.Frank, H.Blocker, and L.Marky, Proc. Natl. Acad. Sci. 83, 3746-55 (1986).
3. M.Fixman and J.Freire Biopolymers 16, 2693-2704 (1977).
4. H.Klump Biochemical Thermodynamics, 100-144 2nd ed. M.Jones ed., Elsevier, NY (1988).
5. F.Crick Cold Spring Harbor Symp. 29, 3-9 (1966).
6. F.Crick, C.Barnett, S.Brenner and B.Watts-Tobin, Nature 192, 1227-32 (1961).
7. M.Nierenberg and M.Matthaei Proc. Natl. Acad. Sci. 47, 1588-1602 (1961).
8. J.Watson et.al. Molecular Biology of the Gene, Benjamin/Cummings 4th ed., pp86-88, Menlo Park. Calif. (1987).
9. H.Klump Free Rad. Res. Comms. 6, 199-200 (1989).
10. J.Marmur and P.Doty, J. Mol. Biol. 5, 109-18 (1962).
11. L.Marky, and K.Breslauer, Biopolymers 21, 2185-94 (1982).
12. V.Boomfield, D.Crothers and I.Tinoco Jr. Physical Chemistry of Nucleic Acids, Harper and Row, NY. (1974).
13. C.Cantor and P.Schimmel Biophysical Chemistry part 1 W.H.Freeman, San Francisco (1980).
14. E.Murrey, and M.Eberle, Nucl.Acid.Res. 17 ,477-97 (1989).
15. P.Sharp, E.Cowe, D.Higgins, D.Shields, K.Wolfe and F.Wright, Nucl. Acid. Res. 16, 8207-8211 (1988).
16. R.Grantham, C.Grautier, M.Gouy, M.Jacobzone and R.Mercier Nucl. Acid. Res. 9, r43-r74 (1981).
17. D.Shields and P.Sharp, Nucl. Acid. Res. 15, 8023-40 (1987).
18. H.Grosjean and W.Fiers Gene 18, 199-209 (1982).
19. H.Pfizinger, P.Guillemaut, J.H.Weil and D.Pillay, Nucl. Acid. Res. 15, 1377-86 (1987).
20. A.Wada and A.Suyama, FEBS Letters 188, 291-294 (1985).
21. A.Wada, H.Tachibana, O.Gotoh and M.Takanashi Nature 263, 439-440 (1976).
22. S.Pedersen EMBO J. 3, 2895-98 (1984).
23. S.Liebhaber, M.Goossens and Y.Kau, Proc. Natl. Acad. Sci. 77, 7054-7058 (1980).
24. D.Cooper, M.Toggart and A.Bird Nucl. Acid. Res. 11, 647-58 (1983).
25. R.Lathe J. Mol. Biol. 183, 1-12 (1985).
26. A.Brownlee Nucl. Acid. Res. 17, 1327-35 (1989).
27. A.Wada and A.Suyama Prog. Biophys. Mol. Biol. 47, 113-157 (1986).
28. M.Sorensen, C.G.Kurland and S.Pedersen J. Mol. Biol. 207, 365-377 (1989).
29. T.Ikemura Mol. Biol. Evol. 2, 13-34 (1985).
30. E.Murrey, J.Lotzer and M.Eberle Nucl. Acid. Res. 17, 477-498 (1989).
31. A.Wada, H.Tachibana, O.Gotoh and M.Takanami, Nature New Biology 263, 439-440 (1976).
32. R.Blake personal. commun.(1990).
33. H.Klump Ber. Bunsenges. Phys. Chem. 91, 206-211 (1987).
34. R.Blake and S.Delcourt Biopolymers 29, 393-405 (1990).