A common structural feature in promoter sequences of *E. coli*

Chang-Shung Tung and Stephen C. Harvey^{1,*}

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 and ¹Department of Biochemistry, University of Alabama, Birmingham, AL 35294, USA

Received October 20, 1986; Revised and Accepted May 20, 1987

ABSTRACT

We have searched promoter regions of *E. coli*, structural genes of the same organism, and computer-generated random sequence DNA for the occurrence of common structural features. This is done by converting the base sequence to a series of numbers representing the sequence of helix twist angles and examining these numerical sequences statistically. Common structural features are shared by the promoter regions with a much higher frequency than are found in structural genes or in random sequences. These structures appear to be scattered randomly throughout the promoters, both in terms of the number of such structures per promoter and in terms of location within each promoter. One particular structure consisting of five successive helix twist angles is reported, along with a list of 60 different hexanucleotide sequences that share this structure. The locations of these structural elements in 61 *E. coli* promoters are also tabulated.

INTRODUCTION

DNA sequences that are involved in the initiation of transcription in *E. coli* have been well characterized (1,2). In addition to possessing common sequences, promoters may contain common structural features that would not be revealed by searches based on sequence homology, because two regions may possess similar structures even though their sequences are not homologous.

This study is aimed at determining what structural features, if any, may be common to the promoters of *E. coli*. Traditionally, a region of DNA is described by specifying the sequence of bases along one strand. Alternatively, it can be described in structural terms by specifying the sequence of the parameters describing the local helix geometry as one moves along the double helix. Such "structural sequences" can then be examined by numerical and statistical procedures to find structural homologies.

Several recent investigations (3-13) have examined the relationships between DNA sequence and the variations in the local geometry of the double helix. The results of these studies have been compared to experimental examinations of the sequence-dependence of DNA bending (14-18) and used to

look for systematic variations of DNA structure in both prokaryotes (19-21) and eukaryotes (22).

We have recently reported the results of conformational energy calculations on the relationship between sequence and structure in DNA, presenting algorithms for the prediction of both helix twist angle and basepair roll angle (9,11). The advantages of this approach are that every atom of each base is explicitly represented and that a large region of conformational space is sampled. This allows the calculation of appropriate thermodynamic averages for the quantities being studied, and it provides information on the deformability and dynamic properties of the double helix, as well as predicting average values for the parameters describing the geometry of the double helix. It is the most recent set of our prediction parameters (11) which we have used here to convert base sequences to sequences of helix twist angles.

It should be emphasized that the structural patterns described here are derived from modeling studies and not from experiment. The specific results are, then, dependent on the accuracy of the model. Since our model (11) does predict helix twist angle patterns that are similar to those described experimentally (3-6), we believe that our fundamental conclusion - that particular patterns of twist angles are found in promoters with a high frequency - is valid. Furthermore, we have a computer program (10) for generating twist angles from all available models (3-6,11), and we will provide the program on request, so this communication is intended to stimulate combined theoretical and experimental searches for homologous structures in a wide variety of DNA sequences.

METHODS

The set of DNA structures to be examined was generated by starting with the base sequences of 61 promoter regions from E. coli, using the Genbank data set (23). For each promoter, a total of 60 bases was taken, from position -55 to position +5, so our data base consists of 3660 basepairs. Each base sequence was converted into a sequence of helix twist angles using a procedure described elsewhere (11).

Our data set of 61 promoters is a subset of those given by Hawley and McClure (24). These 61 promoters were chosen because they all have their sequences known for at least 60 bases upstream from a well-defined transcription start site. For completeness, we subsequently examined all promoters described by Hawley and McClure (24) and tabulated structures in

them. That tabulation is given at the end of the section entitled Results and Discussion.

We decided to look for structural features that cover about six basepairs (five basepair steps), because the most commonly cited consensus sequences in *E. coli* promoters, TTGACA and TATAAT near the -35 and -10 sites, respectively, are both hexanucleotides. In the absence of any other information on potentially significant structures, it seems probable that they would be about the size of these consensus sequences.

A quantitative measure of the difference between two structures is defined by S^2 , the sum of the squares of the deviations in the values of the helix twist angles for the two sequences. Let us consider the comparison of a structure beginning at basepair j in promoter m with a target structure. If we designate the helix twist angle of basepair step n in promoter m by $t(n,m)$ and that of basepair step i in the target structure by $T(i)$, then the relevant value of S^2 is given by

$$S^2 = \sum_{i=0}^4 (t(j+i,m) - T(i))^2 \quad (1).$$

The magnitude of S^2 for stretches of five basepair steps will range from zero (for identical structures) to values greater than 100 deg^2 . All structures were taken to consist of five basepair steps (the upper limit on the sum being $i=4$), as described above. We have found, however, that the results given here are independent of the number of basepair steps over the range $4 \leq N \leq 7$.

One issue to be settled a priori is that of alignment of the structures. In order to broaden the search for interesting structures, we have not required any alignment, and all possible combinations of regions within different promoters have been compared. We have examined the alignment issue in some detail (see Results and Discussion).

A cutoff value of S^2 must be chosen when comparing different structures. Two structures that give a value of S^2 below this cutoff are considered to be similar, while an S^2 value greater than the cutoff value means that the two structures fail to satisfy the similarity criterion. DNA twisting is often modeled by assuming that the torsional energy is proportional to the square of the deformation twist angle (9,11,25; see 9 for references to much of the experimental literature), so S^2 is a measure of deformation energy. We have chosen a cutoff of $S^2 = 5.0 \text{ deg}^2$, corresponding to a deformation energy that is roughly equal to kT , the available thermal energy (9,11,25).

In analogy with traditional sequence-searching methods that identify consensus base sequences, we are looking for the most common sequence of helix twist angles, which we call the consensus structure. To find this structure, we have tested all possible target sequences of six basepairs (all possible target structures of five basepair steps). For each target structure, we have counted the number of structures within the data base that satisfy the similarity criterion. The consensus structure is then the one for which the largest number of similar structures are found.

To test the possibility that the sequences are found at particular locations, we divide the promoter regions, which are each 60 basepairs long, into twelve segments, each of which covers five basepairs. If the structure is randomly located within the promoter, the first base of the structures should be scattered through the first eleven segments with equal probability. (The twelfth segment is forbidden because each structure is six basepairs long.) The appropriate statistic, which compares the observed and theoretical frequencies, is chi square (26).

A second statistical test examines the number of structures found in each promoter. Here we determine $N(0)$, the number of promoters in which no structures are identified; $N(1)$, the number of promoters with one identified structure; $N(2)$, the number with two structures; and so on. Then we ask whether the observed distribution differs significantly from the distribution that would be expected if the structures were distributed randomly over the set of promoters. If the length of each promoter were infinite, the theoretical distribution would be a Poisson distribution. Since we have limited the length of each promoter region to 60 basepairs, we must use a modified binomial distribution that predicts the expected theoretical frequencies, $N^*(0)$, $N^*(1)$, $N^*(2)$, and so on, by counting the ways in which no hexanucleotides, one hexanucleotide, two hexanucleotides, . . . , can be placed into a region 60 basepairs long. The observed and theoretical frequencies are then compared with a standard chi square test (26).

RESULTS AND DISCUSSION

Homologous structures occur in promoter regions with a significantly higher frequency than is observed in other regions of DNA. With a fixed cutoff value of S^2 (5.0 deg^2) and considering hexanucleotide target structures in regions 60 basepairs long, one target structure is found in 58 out of 61 promoters, while no target structures are common to 58 out of 61 regions of structural genes (Table 1). At a lower level of consensus, 55 out of 61

Table 1
Number of target structures found in different classes of DNA at different levels of consensus. Each DNA region is 60 basepairs long, while target structures are 6 basepairs long. The cutoff value of S^2 is 5.0 deg².

Consensus level (Number of sequences in which the tabulated number of structures was found)	Type of DNA region examined			
	61 <i>E. coli</i> promoters (-55 to +5)	61 <i>E. coli</i> genes (+5 to +64)	61 computer generated random sequences ^a	61 computer-generated random sequences ^b
>59	0	0	0	0
58	1	0	0	0
57	4	0	0	0
56	12	1	1	0
55	18	6	2	0
54	27	10	5	0
52	45	24	15	0
50	83	43	37	1
48	137	77	75	8

^a Sequence generation constrained to duplicate observed trinucleotide probability distribution in promoters. Average of two runs.

^b Unconstrained. Average of two runs.

regions, 18 different target structures occur in at least 55 of the promoters, and only 6 target structures show up in at least 55 of the structural gene regions. If we generate artificial random sequences, we find even fewer structures that satisfy the similarity criterion at any level of consensus.

What is the consensus structure within this set of promoters, and what are the sequences that are selected by our search procedure? Let us examine the single structure and the related sequences that are identified when we use a rather stringent similarity criterion, since this is likely to be the most important case. For a structure five basepair steps in length (a hexanucleotide) with a fairly low cutoff value (5.0 deg²), one target structure is found to satisfy the similarity criterion in 58 of the 61 promoters, as described above. This is the consensus structure. It consists of the following sequence of helix twist angles (in degrees): 36.2; 37.8; 26.4; 37.8; 36.2. Since there is often more than one structure within a given promoter that closely matches the consensus structure, a total of 127 structures were found to satisfy the similarity criterion in this case. On eliminating duplicates, 60 distinct sequences were found; these are listed in Table 2, along with the frequency with which each sequence occurs.

There are a number of interesting observations that can be drawn from the data in Table 2. To begin with, we note that these structurally homologous sequences have little sequence homology. For example, the self-complementary sequences AATATT, ATTAAT, and TTTAAA have turned up, as have the sequences TTTGTT (=AACAAA) and GCGGTC (=GACGCC). In ordinary sequence notation, the

Table 2
Sequences from promoters of *E. coli* whose structures have the helix twist angles 36.2°, 37.8°, 26.4°, 37.8°, 36.2° in succession, within a tolerance of $S^2 < 5.0 \text{ deg}^2$.

Sequence	Frequency of occurrence	Sequence	Frequency of occurrence	Sequence	Frequency of occurrence	Sequence	Frequency of occurrence
AACATC	3	AGTGT ^b	1	GGCATC	1	GTAAAT	5
AACATT	1	ATCATC	2	GGCGAA	1	GTGAC	3
AACGCC	1	ATCGTT	1	GGCGAT	1	GTGTGC	2
AACGTC	1	ATTAAT ^a	1	GGCGCT	1	GTGTGT	2
AACGTT ^a	1	ATTATC	1	GGTAAC	2	TTC AAC	3
AATAAC	5	ATTGCT	1	GGTAAT	2	TTC ACT ^b	1
AATAAT	2	ATTGTT	4	GGTGAC ^b	1	TTCATT	1
AATATT ^a	4	GACAAT	3	GGTGAT ^b	1	TTTAAA	4
AATACT	2	GACATC	2	GGTGCC ^b	1	TTTAAT	2
AATATC	1	GACATT	1	GGTGTC ^b	1	TTTATC	2
AATGAT	2	GACGCC	2	GGTGTT ^b	1	TTTATT	8
AATGCC	3	GACGTC ^a	1	GTCAAT	2	TTTGAA	1
AGCGAA	1	GATAAC	2	GTCAAT	1	TTTGCT	2
AGCGTC	1	GATGAA	1	GTAAAA	3	TTTGTC	5
AGTATC	3	GGCAAA	4	GTAAAC ^a	3	TTTGTT	6

Sum of frequencies = 127

^a Self-complementary sequence

^b Contains the "molecular detent" GTG/CAC (ref. 29)

consensus sequence would be NNYRNN, and only the central invariant pyrimidine-purine step would be identified. Although there is almost no sequence homology at the hexanucleotide level, all of these sequences have very similar patterns of helix twist angles, as shown in Table 3. Such sequences have been called structural synonyms (27,28).

We have analyzed the frequency of occurrence of each type of residue at each position (Table 4). These are AT-rich sequences, containing over 70% A+T. As noted above, the central basepairs always form a pyrimidine-purine step. It is interesting to speculate what three-dimensional structure other than a common set of helix twist angles may be common to these sequences: the central pyrimidine-purine step is predicted to bend toward the major groove in some prediction schemes (4-6,11), but in other models (8,12) bending arises primarily from AA steps. Thus, some models (4-6,11) would suggest that there is a bend at the center of these structures, while others (8,12) would say that there are no bends at that location, because of the complete absence of the AA sequence at positions 3 and 4.

Another observation is that, although the consensus structure is characterized by a symmetric pattern of helix twist angles, only six of the 60

Table 3
Detailed structural information for some sequences in *E. coli*
promoters. The consensus structure has the structure listed.

Sequence	Promoter	Location of first base in sequence	Sequence of twist angles (deg)					S ² (deg ²)
AATATT	argC	-7	38.0	38.0	25.8	38.0	35.6	3.9
ATTAAT	trp	-30	34.9	38.2	25.7	38.2	34.9	3.1
TTTAAA	hisB	-53	36.1	38.2	25.7	38.2	36.1	0.7
TTTGTT	galETK	-17	36.1	38.0	26.3	37.9	38.2	4.1
GGCGTC	str operon	-3	35.7	37.7	27.2	38.1	34.5	3.9
Consensus Structure		-	36.2	37.8	26.4	37.8	36.2	0.0

distinct sequences are self-complementary, and they account for only 14 of the 127 total structures. These sequences (and the number of times each one occurs) are AATATT (4), TTTAAA (4), GTTAAC (3), ATTAAT (1), AACGTT (1) and GACGTC (1).

If these structures are biologically relevant, it might be expected that they would tend to occur at particular locations in the promoters, in analogy with the occurrence of the two consensus recognition sequences near -10 and -35 (1,2). We have performed the appropriate statistical analysis (see Methods), and no pattern to the locations of these structures can be inferred.

Table 4
Frequency of occurrence of different nucleotides
at each position in the 127 sequences of Table 2.
Frequencies are expressed as percents.

	Position in hexanucleotide					
	1	2	3	4	5	6
A	33.1	29.9	0.0	63.0	42.5	14.2
C	0.0	0.0	31.5	0.0	13.4	38.6
G	41.7	15.8	0.0	37.0	0.0	0.0
T	25.2	54.3	68.5	0.0	44.1	47.2
Purine	74.8	45.7	0.0	100.0	42.5	14.2
Pyrimidine	25.2	54.3	100.0	0.0	57.5	85.8
A+T	58.3	84.2	68.5	63.0	86.6	61.4

Table 5

Number and locations of 127 structures of Table 2 in 61 *E. coli* promoters. Locations are relative to the transcription start site at +1.

Genbank Promoter Identification	Gene or Protein	Number of structures in promoter	Locations of structures (first basepair)
ECOADA	ada	2	-43 -23
ACOALKA	alkA	2	-29 -21
ECOAMPCFR	ampC	2	-36 -21
ECOARABOP1	araC	1	-14
ECOARABOP2	araBAD	1	-29
ECOARAE	araE	2	-24 -8
ECOARAFGP	araFG	1	-29
ECOARGOP1	argC	4	-54 -49 -13 -7
ECOARGOP2	argE	1	-7
ECOCARAB1	carA (P1)	2	-36 -9
ECOCARAB2	carA (P2)	2	-35 -1
ECOCARAB3	carB	4	-41 -30 -17 -9
ECOCYAP	cya	3	-29 -14 -9
ECODEOC1	deoC (P1)	4	-55 -46 -41 -33
ECODEOC2	deoC (P2)	1	-9
ECODNAAOP	dnaA	1	-16
ECOFOL	fol	2	-54 -3
ECOFUMA	fumA	1	-30
ECOGALE	galETK	3	-44 -17 -3
ECOGALEO	gal	3	-49 -22 -8
ECOGALEP	galE	3	-44 -17 -3
ECOGLNAL	glnL	0	
ECOGND	gnd	1	-29
ECOHISCB	hisB	3	-53 -36 -32
ECOHISOP1	"A"mRNA	2	-29 -25
ECOHISOP2	his	2	-10 -3
ECOILVBPR	ilvB	1	-48
ECOILVGE	ilvGEDA	2	-43 -15
ECOILVIHP	ilvIH	1	-11
ECOLAC	lac (P1)	0	
ECOLEXA	lexA	1	-21
ECOLPP	lpp	1	-40
ECOMALAR	malt	1	-46
ECOMALARA	maltPQ	4	-29 -26 -14 -3
ECOMETAL	metA (P1)	4	-48 -32 -26 -13
ECOMETA2	metA (P2)	3	-54 -43 -21
ECOMETLB1	metBL	3	-49 -41 -14
ECONRDA	nrdA	1	-10

Table 5 (Continued)

Genbank Promoter Identification	Gene or Protein	Number of structures in promoter	Locations of structures (first basepair)
ECOOMPA	ompA	1	-18
ECOOMPC	ompC	2	-48 -29
ECOOMPF	ompF	2	-34 -21
ECOPABB	pabB	5	-42 -36 -21 -15 -8
ECOPFKBK	pfkB	1	-17
ECOPHOA1	phoA	2	-30 -3
ECOPYRBIB1	pyrBI (P1)	1	-16
ECOPYRBIB2	pyrBI (P2)	4	-55 -50 -8 -1
ECORECA	recA	3	-53 -32 -3
ECORHO	rho	1	-2
ECORPLRP01	L11	2	-47 -42
ECORPLRP02	beta operon	2	-24 -18
ECORPSQ	spc operon	2	-45 -8
ECOSTR1	str operon	2	-21 -3
ECOTGMF2	nusA	2	-41 -8
ECOTHRINF	infC	3	-53 -42 -36
ECOTNAA	tnaA	3	-53 -49 -34
ECOTONB	tonB	2	-24 -11
ECOTRP	trp	5	-55 -34 -30 -26 -14
ECOTRPR	trpR	1	-47
ECOUVRA	uvrA	3	-32 -23 -7
ECOUVRC	uvrC	3	-42 -34 -4
ECOUVRD	uvrD	0	

They appear to be randomly scattered over the 60 basepairs upstream from the transcription start site.

How can we explain the apparently random locations of these structures within the promoter in terms of biological activity? Numerous papers have discussed the possibility that local variations in DNA geometry can modulate protein-DNA interactions (3-22,28,29). Perhaps consensus structures serve as pause signals to RNA polymerase. The specific binding of the polymerase to the promoter would then take place because of specific interactions between the protein and the -10 and -35 recognition sequences. In considering this possibility, it is interesting to note that Lu and coworkers (29) have suggested that pausing might occur at the sequence GTG•CAC, which they call a "molecular detent". This trinucleotide is found at the center of several of

the sequences in Table 2. Among other investigations related to this question are those of Nussinov and Lennon (19-22,28), who have used the Calladine-Dickerson rules (4,5) to look for systematic variations in DNA structure in a variety of regulatory sites. They found patterns suggesting that these structural variations might play a role in the recognition and regulation processes in transcription. Whereas Nussinov and Lennon have taken a global view of this problem, looking for regions containing interesting structural patterns, we have taken a more local view, attempting to identify the specific structures (and sequences) that might occur in promoter regions in E. coli.

Another issue we have examined statistically is the number of consensus structures that are found in each promoter. If these structures have a biological function, it would seem likely that every promoter would contain at least one such structure. Structures are indeed found in 58 of the 61 promoters in our data base, a much higher proportion than is found in other DNA sequences or in computer-generated random sequences (Table 1). The appropriate statistical test (see Methods) indicates that the number of structures in each promoter appears to be determined in a random fashion.

How can we explain the apparently random number of these structures that are included in each promoter? Of course, there is an enormous variation among promoters. It is not always possible to positively identify a -35 recognition sequence, and even the Pribnow box at -10, which is almost always present, is subject to significant variations in sequence (1,2). Since we do not know what function these structures might have, we do not know what kind of variability there should be in terms of the number of structures per promoter. The observed distribution may be biologically functional.

The biological significance of these structures might be revealed by an examination of their frequencies and locations within various promoters. To that end, Table 5 lists the promoters in our data base and the locations of all 127 hexanucleotides within those promoters whose structures agree with the target structure within our cutoff criterion. (The full set of sequences is available from either author on request.)

Finally, we have examined those promoters from the data base of Hawley and McClure (24) that were not included in our original analysis, either because a clearly defined transcription start site was not present or because the known sequence upstream from the start site was less than 60 bases long. As expected, many similar structures are found in these promoters (Table 6).

Table 6

Number and locations of structures in those promoters from the database of Hawley and McClure (ref. 24) that are not included in the GenBank database and Table 5.

Gene or protein	Number of structures in promoter	Locations of structures (first basepair)
lacp2	1	10
laci	1	1
malefg	1	24
malk	1	30
aroh	2	32 35
trpp2	4	17 22 31 50
hisa	5	14 17 25 29 49
thr	5	3 11 14 24 36
bioa	2	21 27
biob	1	26
uvrbpl	1	11
uvrbp2	4	6 17 30 44
uvrbp3	2	4 23
porir	2	32 49
poril	1	3
spot42rna	1	26
mlrna	0	
alas	1	10
trps	2	39 53
glns	1	14
tufb	1	55
tyrt	0	
leultrna	3	3 7 17
supbe	3	14 17 53
rrnabp1	2	2 27
rrng p1	1	27
rrnd p1	1	2
rrne p1	1	48
rrnx p1	1	48
rrnab p2	1	6
rrng p2	0	
rrndex	1	14
sl0	1	5
rpoa	1	49
rplj	2	26 32
rpob	1	45

In summary, we have shown that promoters in E. coli contain homologous structural regions with a frequency that is substantially greater than that with which homologous structures are found in structural genes or random

sequence DNA. It has often been argued (3-22,28,29) that particular kinds of DNA structures may play a role in modulating the interactions between proteins and DNA. Since hard evidence for this belief is still lacking, it is important to identify specific structures and the set of associated sequences that might be involved in such modulation. We have identified one candidate consensus structure and a large number of sequences that share this structure (Table 2), in the hope that this will assist in the search for experimental evidence that proteins recognize specific structural features.

ACKNOWLEDGMENTS

We thank Drs. Jeffrey D. Esko, N. Patrick Higgins, Charles T. Turnbough, Jr., and Robert D. Wells for critically examining a preliminary version of this paper and for many suggestions. This research was supported by the Department of Energy and grants from the National Institutes of Health (GM-34015 to SCH and GM-26857 to G.C. Salzman).

*To whom correspondence should be addressed

REFERENCES

1. von Hippel, P.H., Bear, D.G., Morgan, W.D., & McSwiggen, J.A. (1984) Ann. Rev. Biochem. **53**, 389-446.
2. McClure, W.R. (1985) Ann. Rev. Biochem. **54**, 171-204.
3. Kabsch, W., Sander, C., & Trifonov, E.N. (1982) Nucl. Acids Res. **10**, 1097-1104.
4. Dickerson, R.E. (1983) J. Mol. Biol. **166**, 419-441.
5. Calladine, C.R. (1982) J. Mol. Biol. **161**, 343-352.
6. Fratini, A.V., Kopka, M.L., Drew, H.R., & Dickerson, R.E. (1982) J. Biol. Chem. **257**, 14686-14707.
7. Haran, T.E., Berkovich-Yellin, Z., & Shakked, Z. (1984) J. Biomolec. Struct. Dynamics **2**, 397-411.
8. Prunell, A., Goulet, I., Jacob, Y., & Goutorbe, F. (1984) Eur. J. Biochem. **138**, 253-257.
9. Tung, C.-S., & Harvey, S.C. (1984) Nucl. Acids Res. **12**, 3343-3356.
10. Tung, C.-S., & Harvey, S.C. (1986) Nucl. Acids Res. **14**, 381-387.
11. Tung, C.-S., & Harvey, S.C. (1986) J. Biol. Chem. **261**, 3700-3709.
12. Trifonov, E.N. (1985) CRC Crit. Revs. Biochem. **19**, 89-106.
13. Zhurkin, V.B., Lysov, Y.P. & Ivanov, V.L. (1979) Nucl. Acids Res. **6**, 1081-1096.
14. Marini, J.C., Levene, S.D., Crothers, D.M., & Englund, P.T. (1982) Proc. Natl. Acad. Sci. USA **79**, 7664-7668.
15. Wu, H.-M., & Crothers, D.M. (1984) Nature **308**, 509-513.
16. Hagerman, P.J. (1984) Proc. Natl. Acad. Sci. USA **81**, 4632-4636.
17. Koo, H.-S., Wu, H.-M., & Crothers, D.M. (1986) Nature **320**, 501-506.
18. Hagerman, P.J. (1986) Nature **321**, 449-450.
19. Nussinov, R., & Lennon, G.G. (1984) J. Biomolec. Struct. Dynamics **2**, 387-395.
20. Nussinov, R. (1985) J. Theor. Biol. **115**, 179-189.
21. Nussinov, R. (1984) J. Biol. Chem. **259**, 6798-6805.
22. Nussinov, R., & Lennon, G.G. (1984) J. Mol. Evol. **20**, 106-110.

23. Burks, C., Fickett, J.W., Goad, W.B., Kanehisa, M., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C.-S., & Bilofsky, H.S. (1985) Comp. Applic. Biosci. **1**, 225-233.
24. Hawley, D.K., & McClure, W.R. (1983) Nucl. Acids Res. **11**, 2237-2255.
25. Tidor, B. Irikura, K.K., Brooks, B.R., & Karplus, M. (1983) J. Biomolec. Struct. Dynamics **1**, 231-252.
26. Dixon, W.J., & Massey, F.J., Jr. (1969), "Introduction to Statistical Analysis" (McGraw-Hill, New York), p. 238.
27. Beyer, W.A., Burks, C., & Goad, W.B. (Fall, 1983) Los Alamos Science **9**, 62-63.
28. Lennon, G.G., & Nussinov, R. (1984) J. Mol. Biol. **175**, 425-430.
29. Lu, P., Cheung, S., & Arndt, K. (1983) J. Biomolec. Struct. Dynamics **1**, 509-521.