RC 20804 (92167) 21APR97
Bioinformatics / Data Mining 11 pages

IBM Research Report

Preliminary Results On The Discovery Of Patterns Of Amino Acids Common To Sequences Of Core Histones 3 & 4

Isidore Rigoutsos Aris Floratos
IBM Research Division
T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Christos Ouzounis
EMBL Cambridge
The European Bioinformatics Institute
Hinxton Genome Campus
Cambridge CB10 1SD, England, UK

LIMITED DISTRIBUTION NOTICE
This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g. payment of royalties).

IBM Research Division
Almaden . TJ Watson . Tokyo . Zurich . Austin

Preliminary Results On The Discovery Of Patterns Of Amino Acids Common To Sequences Of Core Histones 3 & 4

Isidore Rigoutsos[1,2] *Aris Floratos*[1,3] *Christos Ouzounis*[4]

March 1997

Abstract

Pattern discovery (an NP-hard problem), in contrast to pattern matching (a problem solvable in polynomial time), is a challenging problem. Using *Teiresias*, a newly developed algorithm, we have explored (see [5]) a number of test cases in order to identify patterns of amino acids that appear across sequences (common motifs) as well as within individual sequences (internal repeats). In that work, only a small subset of the available results were presented in order to showcase the ability of the algorithm to: (a) validate the approach through the discovery of previously reported patterns; (b) demonstrate the capability to automatically identify highly selective patterns particular to the sequences under consideration; and (c) discover unidentified patterns in the well-studied example cases used. One of those example cases was the leghemoglobin family of proteins. In this report, we present the full range of results obtained for this particular protein family.

1. Introduction

We herein present preliminary results obtained with the *Teiresias* pattern-discovery algorithm using the amino acid sequences of the proteins known as plant hemoglobins.

Teiresias [1], is a novel pattern discovery algorithm that has the ability when given a set of event streams in any alphabet of N+1 characters (N letters + space) to discover *all* the patterns occurring at least twice in *any relative arrangement* within the set of event streams.[5,6] What this means is that the algorithm will discover *anything* that appears *at least* twice (or, any desired number of times greater than 2). By "anything" here we mean that we impose no restriction on the number of characters that will participate in the pattern; furthermore, we impose no restriction on the relative arrangement that the characters will have.

Let us illustrate the above with a specific example. Figure 1 shows an event stream of 2,000 characters from the English alphabet. In this stream, an adversary has hidden a phrase several times and has told us nothing as to the length or the composition of the phrase. The only thing that we have been told is that the phrase was hidden a minimum of two times.

A simple, straightforward approach would be to exhaustively generate first all strings of length l_{min} and check them against every possible location of the stream, then all strings of length l_{min}+1, all strings of length l_{min}+2, and so on. If the stream contained N=2000 characters, the minimum length was l_{min}=10, and $|\Sigma|$=27 (26 letters plus space), then ignoring boundary conditions, just the first iteration would require $N\,|\Sigma|^{l_{min}} \sim 5.4^{13}$ operations! For the shown stream, the embedded phrase has a length of 34, making such an approach prohibitive.

Teiresias can however determine the phase and the locations where it appears in a matter of seconds on an IBM Power-PC workstation using only a few Megabytes of main memory. The hidden phrase is

---

[1] IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA.

[2] Corresponding Author.

[3] Also: Courant Institute Of Mathematical Sciences, New York University, New York, NY 10012, USA.

[4] EMBL Cambridge, The European Bioinformatics Institute, Hinxton Genome Campus, Cambridge CB10 1SD, England, UK.

[5] Note that there can also be a single event stream at the input, in which case *Teiresias* will identify what is typically referred to as *internal repeats*.

[6] A pattern is defined to be any combination of characters (here, amino acids) and "don't care" characters with any spatial extent, which appears at least twice in the input under consideration.

"F.A..S.T...AN.D..A.C...C.U.R..A.TE" (where the dots are occupied by characters of the otherwise random stream) and appears a total of five times in the shown stream.

```
>example 1
QOJPYBHHLSJUEXDFNYHLDSUXGTDLYHTUBLOCNDCXXIGUYQLQGLYFQJWGFPUTKVRYFNNJQXEQXKJYKAMNONRFOA
OYSGTOLWANADTLAFCIMFCMUPRMRATTEGEZSMZFANOTCTVYDAQPJXMSNQQZKUTFXLDYMXOCIGIZRMCYPKVTVBMN
LXYOHNDRCKFHEVGUQQEHQTQGFCXJDYACDJXKSSTFHAMUSBTJEQANMDICACCLXMCSUARASAGTEXVTVIEZEOAQWF
DAFPSITLOKANEDXOAJCWLNCVUBRCSAOTEIBHAUDOPHXFIZNBFMJRRHGYPPZGUNDKKUHJGSBHPIXYLBAXMPLOCG
DVKXTSGSDXDPNZHWIPAZHEDWEVFPJNSNSDBLHXRCJCGJXHTNIQCQKOYFFNKZMPNHGKCVGCBCOOACSQWXJFRDJY
CINMPLUTNSXGKMOBZVKDHTMXXWJGXVJGCYWLDDXLIOPDPBUXEJNBPZDKDPPVYXPHFTHYXRUPCXVVXIQDWMGDTM
WBMBDDXTPCZBXJPWMZJRMKEFFYADPSKTTSFANCDQXAICHJHCOUHRGBALTEOKCZZXETQMUOKFHAWPSATMWGANED
OIAMCVFMCCUZRWSAJTEBKBYUOEEVPWQXAIXXIESOKKLXJTZUDEUIFSHTHZYFVIFJLPYCLZKYTWRWTAWKYYPNRI
EAARSLIZNWWCFORKISDYYCCESGJTYWPYUWRWUEMLWXZQDLBZVJMIOOSQAIJDETDAHVRFPZKWOAJHALBMSBCMNI
PILEDYEBTXELJJWEFCEGTXGIHMJOOSCKFGIRRRYOOUZNNAJOZDKHLVTTOCYGWVILKQKDLGRSKRMIBUBHETTDDO
TWFGKTYEVNZVTTLCLZBCOKCQDTDPJJSVSZIZQRBFPYOKBCMWSPPAAPQPYJBLAKHWVNPYYPMPFIEKRNXINYIQKC
SZVYQHUWCZSNTFWXXNDNVLILVCRFOYDMRUPOAOPDFCSLVSRCV
```

Figure 1. An event stream of 2,000 characters which contains a hidden phrase of unknown length and composition at least two times. Each circle denotes the beginning of an occurrence of the hidden phrase.

Even knowing that the hidden phrase begins at the locations marked with the circles (see Figure 1), it is not easy to discover what the phrase is or how long it is. Incidentally, "F.A..S.T...AN.D..A.C...C.U.R..A.TE" is the only collection of letters (and spaces), of *any* length that appears 4 or more times in the stream.

Although we have demonstrated the nature of the algorithm using a single stream of letters, the method directly extends to the case where the input is a collection of several sequences of letters. Also, the algorithm imposes no restriction on the size of the alphabet that the sequences can be built from. We fact in fact tried the algorithm on input sequences built from an alphabet of 90 characters with the same high efficiency.

One of the things contributing to the algorithm's strength is the loose definition of what a pattern is. There is no restriction on the symbol composition of the pattern, its spatial extent, minimum length, maximum length or minimum number of occurrences.

Moreover, the algorithm does not make use of any alignment to discover the patterns. Alignment is typically unable to handle domain interchanges. For example, if *a*, *b*, and *c* denote three patterns, it may happen that several of the input sequences will contain the patterns in the order *a-b-c*, a second set will contain the patterns in the order *b-a-c* and a third set will contain the patterns in the order *c-b-a*. Any approach that relies on alignment is bound to perform poorly on such a data set; and this will be particularly true if the three occurring patterns contain only a few amino acids while at the same time spanning a large portion of the sequence (=weak patterns).

Another characteristic of the algorithm is that, although it is deterministic, it does not require any search of the corresponding search space; although counterintuitive, such a thing is possible and contributes to the algorithm's speed. And, of course, in this manner, the algorithm requires no clustering or classification steps and thus avoids carrying out any statistical evaluation of the input sequences.

The rest of the presentation contains the full set of results obtained by applying Teiresias on the protein families known as core Histones 3 and 4.

2. Results

Core histones have been suspected to be homologous for a long time [2], but it was only recently that detailed analyses established this evolutionary relationship [3]. All histones and the eukaryotic transcription factors CBF-A/C have been found to contain sequence patterns that are characteristic of the core histone fold [3]. These

patterns contain, among other motifs, the pentapeptide KRKT[IV][7], towards the end of the C-terminal domain [3]. To demonstrate the power of the Teiresias algorithm, we have analyzed the core histone families H3 and H4 and compared the results to those determined previously.

For the purpose of carrying out the experiments, we generated an input set containing 19 sequences of the H3 family and 18 sequences of the H4 family. The PIR Rel. 47 [6] names of the resulting set appear in the table below.

HSHU33	HSCH3	HSRK3	HSTR3	HSFI3	HSXL31	HSUR3M	HSUR3P
HSBO3	HSEAH3	HSKW3	HSPM3	HSBH3[8]	HSXL32	HSZP3	HSBY3
HSVK3L	HSDK34	HSMS34	HSHU4	HSBO4	HSPG4	HSRT4	HSCH4
HSTR4	HSXL4	HSUR4	HSUR4P	HSFF4	HSZM4	HSWT4	HSPM4
HSWT41	HSBY4	HSZP4	HSTE42	HSTE41			

Table 1. The PIR labels of the 37 sequences from the H3 and H4 families that were used in the experiments.

Processing the input set with *Teiresias* took place in a matter of a few minutes on an IBM Power-PC workstation and a large set of patterns was discovered. The discovered set comprised patterns that were common to both families as well as patterns that were common to only the members of a single family.

The pattern T...I.....V..I.....R which is the only one appearing in all of the H3 and H4 family members of the input set can be found in 61 entries in the latest Non-Redundant Database (effective date 26 February 97 containing approximately over 62 million amino acids). Of those 61 entries, 51 are histones, whereas the remaining 10 proteins include:

P15699	AMDR_EMENI ACETAMIDASE REGULATORY PROTEIN
Q06943	HMGZ_DROME HIGH MOBILITY GROUP PROTEIN Z
P36733	VL1_HPV12 MAJOR CAPSID PROTEIN L1
P50790	VL1_HPV24 MAJOR CAPSID PROTEIN L1
Q06895	VP4_ROTHY OUTER CAPSID PROTEIN VP4

four more hypothetical proteins, and a mutant protein sequence from PDB (1MAS-B). Ironically enough, HMG-Z belongs to the so-called *non-histone* chromosomal proteins and may represent a false positive. Yet, the pattern T...I.....V..I.....R is very specific, and identifies *all* histones in the Non-Redundant Database (many more than the original set on which *Teiresias* operated). Furthermore, the pattern is present in proteins which are either DNA-binding, viral, or hypothetical and thus in essence represents predictions of either evolutionary or functional similarity.

The following table contains the patterns discovered by Teiresias when run on the set of sequences described in Table 1. This set includes all the patterns that appear in at least 18 of the 37 input sequences.

In order to be able to assign a quantitative measure to the importance of the reported results we employed the following simple scheme. First, we determined for each of the 20 amino acids their probability of naturally appearing. Then, we assumed a model of independence and identical distribution for the random variable denoting the probability that a specific amino acid participates in a given pattern: a probability of appearance was assigned to each of the discovered patterns as the product of the respective probabilities.

Clearly, for a given set of input sequences, this scheme could favor a long pattern that appears 2 times in the entire data set over a much shorter pattern that appears 10 times. We chose to alleviate this problem by sorting in order of decreasing expected importance only those patterns which appear the same number of times in the input dataset: this modification will correctly classify a longer pattern as more important than a shorter one given that they both appear equaly frequently in the dataset.

---

[7] The notation [IV] here is used to indicate that the specified position can be occupied by either of the amino acids I, V.

[8] This is an 80 amino-acid fragment.

Each pattern is represented by an entry in the table that follows. Every such entry has the following format:

<Multiplicity #> <# Of Sequences> <LogProb> <Pattern>

where

- *<Multiplicity #>*: the number of times that the given pattern appears in the input sequences (including internal repeats).
- *<# Of Sequences>:* the number of distinct sequences the given pattern appears in.
- *<LogProb>:* the logarithm of the probability associated with the given pattern.
- *<Pattern>:* the pattern itself.

37 37 -14.597255 T...I.....V..I.....R	29 19 -9.921900 S...L.....L....A
37 37 -8.832016 I....K...R	19 19 -9.631621 H.K.....R
37 37 -8.453389 AI.....K	19 19 -9.554899 K.....K.H
37 37 -8.346226 I.....RL	19 19 -9.554899 K.....K.....H
37 37 -8.346226 L.....R....I	38 19 -9.516501 FQ.....Q
36 36 -16.994173 T...I..L.....V..I.....R	20 19 -9.360166 A.....IH
36 36 -16.781574 E.....V...E.....V....K.....G	19 19 -9.252994 A..H.K
36 36 -16.345238 K.A.....GGVK	19 19 -9.135627 R.....N...I
36 36 -13.727928 V.....T.....K.V.....L	21 19 -9.126381 Q....E.....Q
36 36 -11.443268 AI.....K.I	19 19 -9.020659 R.A.....Y
36 36 -11.305656 I.....RL.R	19 19 -8.996786 VR.....F
37 36 -11.275206 L.RR....R	19 19 -8.939188 I....I.....R
36 36 -11.228934 I.K....RL	36 19 -8.908738 RR.....I
36 36 -10.743144 L.....R....I.....L	19 19 -8.883363 K.....D.....R
36 36 -9.871938 D.....H.....D	19 19 -8.868488 R.....D..T
36 36 -8.908738 I.....RR	19 19 -8.832016 I.....K.....R
53 36 -8.908738 I.....R..R	19 19 -8.832016 I....K.....R
36 36 -8.905188 V.....T...F	19 19 -8.801566 K.....R..R
36 36 -8.885229 KT.....Q	35 19 -8.801566 RK.....R
47 36 -8.883363 KR.....D	34 19 -8.801566 R.....K...R
37 36 -8.832016 K.....I.....R	34 19 -8.801566 R.....KR
36 36 -8.618159 F...V...A	30 19 -8.801566 R.....K....R
36 36 -8.572834 D....L..Q	19 19 -8.801566 K...R..R
36 36 -8.553702 I.....R..G	37 19 -8.790891 V.....D.....R
36 36 -8.530111 A.....I.....R	19 19 -8.790891 V.....D.R
36 36 -8.446530 K.....RG	37 19 -8.743664 P.....EI
53 36 -8.315776 L.R.....R	19 19 -8.739544 V.....I...R
52 36 -8.239054 K.....L....R	19 19 -8.724844 K...RK
36 36 -8.162332 K....L.....K	20 19 -8.724844 K.....K....R
36 36 -8.136626 K..L.....E	19 19 -8.724844 K.....K.....R
52 36 -8.091493 GG.K	34 19 -8.724844 K.....RK
53 36 -8.091493 K.....GG	35 19 -8.709968 T.....K.....R
36 36 -8.054985 L.....VT	19 19 -8.709968 TK.....R
36 36 -7.937149 LA.....R	19 19 -8.709968 T.....RK
40 36 -7.834721 L.....E.A	19 19 -8.709968 T...K.....R
36 36 -7.765997 A....A...G	36 19 -8.709968 K.T.....R
35 35 -19.646786 V.....T.....K.V.....L.R..R	36 19 -8.684262 T.....R...E
35 35 -17.832678 L.....D.....H..R.....D.....G	31 19 -8.674921 Q.....R....A
35 35 -17.556685 T...IR.....V..I.....R	32 19 -8.674921 Q..R.....A
35 35 -16.595759 V.....T.....K.VT.....L	21 19 -8.648122 K...K.....K
35 35 -15.344163 D.....H....T....D.....G	19 19 -8.648122 K...K....K
35 35 -14.188363 I.K....RL.R	20 19 -8.648122 K..K.....K
35 35 -13.856008 LARR....R	37 19 -8.648122 K...K.....K
35 35 -12.914869 L....A.....LR.....A	19 19 -8.633246 T.....K...K
35 35 -12.452740 D.....HA.....D	31 19 -8.632372 K...R.....V
35 35 -11.543581 I....I....R..G	19 19 -8.632372 K.....R...V
35 35 -11.436409 G..K...R.....I	29 19 -8.606666 R.....E.....V
35 35 -11.228934 L...I....K...R	19 19 -8.606666 V.....ER
35 35 -11.027332 A.RK.....G	19 19 -8.584618 V.....F.....S
35 35 -10.743144 I.....RL.....L	20 19 -8.572834 L.....Q.....D
35 35 -10.743134 K.....LA.....K	20 19 -8.572493 Q....A.....E
35 35 -10.415523 L.....E.A.....A	34 19 -8.554850 T.....P....V
35 35 -9.083998 IQ.....R	37 19 -8.505727 Q.....A.....V
35 35 -8.894358 E.....F...V	19 19 -8.499661 R.A.....R
50 35 -8.878288 R.....RR	32 19 -8.499661 A.....R.R
37 35 -8.832016 I.K.....R	19 19 -8.499661 R.A.R
71 35 -8.801566 K.....R..R	20 19 -8.489860 A.....DT
36 35 -8.801566 K....R..R	19 19 -8.476980 I.K.....G
35 35 -8.756999 A.....N...I	20 19 -8.466121 R.....S.....R
35 35 -8.738920 RK.....P	19 19 -8.463178 V.....V.K
35 35 -8.709968 TK...R	36 19 -8.446530 R.....K.....G
50 35 -8.617496 T....R.....V	19 19 -8.427683 I.....E.A
50 35 -8.581458 A.R.....D	20 19 -8.422939 A.R...K
35 35 -8.529944 E....K.....V	20 19 -8.422939 K...R.....A
36 35 -8.525899 T.....VT	36 19 -8.422939 A.....K.....R
35 35 -8.523252 R....G...R	35 19 -8.422939 RK.....A
53 35 -8.505727 Q.....A.....V	35 19 -8.414315 K.....QL
36 35 -8.499661 A.R.....R	36 19 -8.408725 A.....F.....A
51 35 -8.499661 AR.....R	35 19 -8.408063 T.....A.R
53 35 -8.499661 A.R....R	19 19 -8.408063 T....R.A
41 35 -8.499661 ARR	36 19 -8.399439 T..L.....Q
36 35 -8.448303 V.....VT	22 19 -8.399439 T.L.....Q
35 35 -8.448303 V.....TV	20 19 -8.397233 R....A.....E
50 35 -8.422939 A.RK	34 19 -8.397233 R...E.A
52 35 -8.422939 A..K.....R	21 19 -8.394143 S.E.....I
46 35 -8.422939 AKR	21 19 -8.389398 KR.S
35 35 -8.383884 G..KP	20 19 -8.389398 R...KS
53 35 -8.369808 GK...K	33 19 -8.389398 KS.....R
52 35 -8.369808 G.K...K	37 19 -8.388609 Q.....L....E
36 35 -8.346217 K.....A.....K	19 19 -8.369808 K.....G.....K
35 35 -8.346217 K...K..A	22 19 -8.369808 G....K.....K

35 35 -8.331341 T..A.....K	53 19 -8.360293 K.P.....A
35 35 -8.320852 L.....K.....D	33 19 -8.354058 G.V.....R
37 35 -8.269504 K.....I.....L	19 19 -8.346226 RL.....I
37 35 -8.253745 A..K.....V	19 19 -8.346226 I.....L..R
35 35 -8.253745 V.K.....A	19 19 -8.346217 K.....AK
35 35 -8.253130 P....L.....R	38 19 -8.331341 T...A.....K
49 35 -8.224179 T.....L....R	19 19 -8.330467 A....RV
35 35 -8.162332 K.....K.....L	33 19 -8.319885 Q.A....G
35 35 -8.136626 E.....K..L	33 19 -8.319885 Q.....G..A
35 35 -8.087493 A.....R.S	33 19 -8.316466 T...A.....T
35 35 -8.087493 AR.....S	20 19 -8.315776 RL.....R
35 35 -7.834721 L....E.....A	19 19 -8.312676 K....KS
35 35 -7.791546 G.V.....L	34 19 -8.305635 T.....E.A
34 34 -24.943437 K..A.....Q...K.A.....GGVK	36 19 -8.297801 KST
34 34 -22.514618 V.....T.....K.VT.....L.R..R	52 19 -8.296294 Q.A.....A
34 34 -20.700510 L.....D.....H..R.T....D.....G	19 19 -8.277336 V...K.....G
34 34 -20.413480 L.....D.....HA.R.....D.....G	20 19 -8.254629 T..L.....I
34 34 -19.953602 T...IR.L.....V..I.....R	37 19 -8.243798 E.....L.....I
34 34 -17.924965 D.....HA...T....D.....G	20 19 -8.239054 K...R.....L
34 34 -17.924965 D.....H....T....D...A....G	19 19 -8.228039 A....V.....E
34 34 -14.265085 L.....R..I.....RR	19 19 -8.224179 L....R...T
34 34 -14.171989 P....L.RR....R	19 19 -8.213348 R.....L..E
34 34 -14.111641 K.....L...I....K...R	19 19 -8.213348 R.....LE
34 34 -14.018545 K..RK.L.....P	19 19 -8.198666 G.....G.....I
34 34 -13.702573 L.R.....R....I.....L	35 19 -8.176408 K.....P....L
34 34 -13.702573 I.....RL.R.....L	30 19 -8.144625 G.....A.R
34 34 -13.625851 I.K....RL.....L	19 19 -8.144625 G.....A.R
35 34 -11.791445 K.....I.....R..R	19 19 -8.144625 A.....G.....R
34 34 -11.475160 E.....F...V...A	37 19 -8.121751 L.E.....T
34 34 -11.331010 V.....TV.....K	33 19 -8.121033 A..A.R
34 34 -11.280281 L.....KR.....D	36 19 -8.121033 R.....A...A
34 34 -11.213174 A..K.....V....R	35 19 -8.112410 Q.....A....L
34 34 -11.153636 A.D....L..Q	19 19 -8.091493 G.....K.G
34 34 -11.046923 AR.....R.S	19 19 -8.067902 G.K.....A
34 34 -11.019333 K.....E.....K..L	19 19 -8.067902 A.....K.....G
34 34 -10.993628 E.....K..L.....E	19 19 -8.067902 A..K.....G
34 34 -10.974201 GGK...K	19 19 -8.067902 G....K.....A
34 34 -10.950610 GK...K..A	34 19 -8.044311 K.....A...A
34 34 -10.950610 G.K...K..A	37 19 -8.044311 A...K.....A
34 34 -10.901654 L....AK.....D	19 19 -8.044311 A.....A..K
34 34 -10.850306 L.....AI.....K	19 19 -8.044311 A..A.....K
34 34 -10.833932 P....LA.....R	37 19 -8.044155 V.L.E
34 34 -9.662071 H.K.....I	36 19 -8.018605 A.....EA
34 34 -9.276585 G..K.H	21 19 -7.967599 I.....L....A
34 34 -9.230092 R...V.....Y	19 19 -7.960740 R.L.G
34 34 -8.976826 Q...K.....R	36 19 -7.951840 A.....V.....A
35 34 -8.908738 IRR	36 19 -7.951840 A..V.....A
34 34 -8.833555 E...V.....N	35 19 -7.951840 A..A..V
34 34 -8.832016 K.....IR	20 19 -7.937149 R....A.....L
34 34 -8.832016 K.....I.R	19 19 -7.903609 R..S.....L
34 34 -8.815642 R.....P..R	37 19 -7.874503 P...L....A
35 34 -8.807633 Q...K.....V	36 19 -7.860427 K.....LA
48 34 -8.792757 Q..T.....V	19 19 -7.860427 K.....L....A
34 34 -8.739544 V.....I....R	20 19 -7.860427 A...K.....L
49 34 -8.724844 K..RK	19 19 -7.860427 L....A.K
34 34 -8.724844 K.....K.....R	19 19 -7.858312 G.L....E
34 34 -8.709968 R...T.....K	20 19 -7.845552 A....T.L
34 34 -8.606666 E...R.....V	22 19 -7.845552 T.L....A
34 34 -8.583324 T.....A...Q	20 19 -7.835062 D....L.....L
34 34 -8.515069 V...E.....T	38 19 -7.835062 L.....D..L
34 34 -8.515069 TE.....V	21 19 -7.834721 A....L.E
45 34 -8.504736 AK.....D	37 19 -7.834721 A.....E.....L
34 34 -8.476980 G..K.....I	38 19 -7.834721 AL.E
52 34 -8.446530 G..K....R	20 19 -7.834721 E.....L.....A
34 34 -8.431654 R...T.....G	19 19 -7.813179 GG.....G
51 34 -8.422939 K.....A.R	35 19 -7.783715 L..I.....L
35 34 -8.383884 K.....G...P	20 19 -7.783715 L.I.....L
34 34 -8.374523 T.....R.....S	32 19 -7.742406 A..A.....A
50 34 -8.369808 G.....K...K	38 19 -7.742406 A.....A..A
34 34 -8.315776 R.....R.....L	35 19 -7.734415 S.....LV
34 34 -8.305635 A.....E.T	38 19 -7.661667 T.L.....L
34 34 -8.294805 E.....E.....A	37 19 -7.661667 L.....T.L
34 34 -8.239054 RK.....L	19 19 -7.584071 LV.....L
34 34 -8.147457 T.....K.....L	20 19 -7.584071 L.....LV
34 34 -8.121751 L....E.....T	31 19 -7.582113 G.....L....A
34 34 -8.121033 A.R.....A	48 19 -7.558522 A....L.....A
34 34 -8.110920 E.....L....E	20 19 -7.558522 A....L.....A
36 34 -8.091493 GGK	29 19 -7.524982 L...S.A
34 34 -8.067902 GK.....A	19 19 -7.374638 A.....L.....L
48 34 -8.067902 G..K.....A	36 19 -7.374638 L.....L....A
34 34 -7.951840 A.....A..V	18 18 -168.91385 KSTELLI...PFQRLVREIAQDFKT.LRFQ..A..ALQE..EAYLV.LFEDTNL.AIH.KRV.I..KD..L .RR.RGER
37 34 -7.937149 R.....L....A	18 18 -147.93743 STELLI...PFQRLV.EIAQDFKT.LRFQ..A..ALQE..EA..V.LFEDTNL.AIHAK.V....KDI.L. ....GER
34 34 -7.860427 K.L....A	18 18 -147.81544 STELLI...PFQRLV.EIAQDFKT.LRFQ..A..ALQE..EA..V.LFEDTNL.AIH.K.VT...KD..LA ....GER
35 34 -7.860427 L....AK	18 18 -147.64632 STELLI.KLPFQRLV.EIAQDFKT.LRFQ..A..ALQE..EA..V.LFEDTNL.AIH.K.V....KD..L. ....GER
34 34 -7.860427 K.....AL	18 18 -144.91402 STELLI...PFQRLV.EIAQDFKT.LRFQS.A..ALQE..EA..V.LFEDTNL.AIH.K.V....KD..L. ....GER
51 34 -7.834721 E.....L.....A	18 18 -100.68136 A.TKQTA.KST..KAPRKQL..KAA.K.AP..GGVKK.H...P.TVAL.EI.....L
34 34 -7.783715 LI.....L	18 18 -99.060001 G.....I.GITKPAIRRLARRGGVKRIS...Y...R.VLK.FLE.V.RD.VT
34 34 -7.690619 P....L.....L	18 18 -33.882511 KGGKG.GK.GAKRH
34 34 -7.676543 K....L.....L	18 18 -25.945293 Q..T...IR.L.....V..I.....E.R
33 33 -23.281312 L.....D.....HA.R.T....D.....G	18 18 -25.626430 K.....K.....L...I....K...R.....G....S...Y
33 33 -23.281312 L.....D.....H..R.T....D...A....G	18 18 -25.212236 A..KQ.A.K.....AP....V.K
33 33 -20.505767 D.....HA...T....D...A....G	18 18 -23.696295 L....A.....LR.....I.....A...G.....L.....V
33 33 -19.362376 E.....V...E.....A..V....K.....G	19 18 -22.631434 A..KQ...K.....AP....V.K
33 33 -19.213070 TK.A.....GGVK	18 18 -22.611302 K.....A.R.....K.....I.....RL.R
33 33 -17.731945 Q..T...I.....V..I.....R	18 18 -22.512628 G..G.G.G.....R.....N...I.....R
33 33 -16.994348 K.....K.....L...I....K...R	35 18 -22.422001 A..KQ.A.K.....AP.....K

33 33 -16.752791 P....LARR....R	18 18 -21.652288 A..K.....G.V...........A..E.........L..R
33 33 -16.622938 G.K..RK.L.........P	18 18 -20.715200 D......HA...TV...D........G
33 33 -16.585281 I.K....RL.R.........L	18 18 -20.675861 K........K......P........HR....T..L
33 33 -15.874298 R......L....A......LR..............A	29 18 -20.275206 K........K......P.T........Y...T..L
33 33 -15.580930 H.K........I......R..R	18 18 -20.035400 T...I..L.....V..I......D.R
33 33 -14.265085 R.........I......RL.R	18 18 -20.030499 RK........I.K....RL.R
33 33 -14.231403 R.G.V........Y......L	18 18 -19.934286 GG.KR......D......K.......R
33 33 -13.876335 K........E.....K..L...........E	36 18 -19.841199 A..KQ...K.....AP......K
33 33 -13.861083 L....AKR......D	18 18 -19.806217 G....GK...K.......K.....I.........R
33 33 -13.840186 L..........AI.........K.I	18 18 -19.784658 K........K......P.T............T..L............I
33 33 -13.833317 K...K..A............G..K	32 18 -19.754208 K........K......P.T........R....T..L
33 33 -13.595216 L.........R.V........T.......A	18 18 -19.741003 R.......E......V...E..........V....K.........G
33 33 -13.555003 GGK...K..A	18 18 -19.651872 A........K.....I.K....RL.R
33 33 -12.236014 G..K.HR	18 18 -19.649406 E......V...E...........V....K....T....G
33 33 -11.791445 R..K.....I.........R	18 18 -19.452238 L.RR....R....I.......L.......V
33 33 -11.791445 K..R..........IR	18 18 -19.364760 R..P......EI.....T.L.........L
33 33 -11.791445 K............IRR	18 18 -19.316427 A........K.....I......RL.R..............S
33 33 -11.566095 R......E...R........V	18 18 -19.247372 K........E.....K..L...........E.A....V
33 33 -11.489373 E...RK......V	18 18 -18.161732 A..H.K........I......R..R
44 33 -11.464165 AKR.......D	18 18 -18.007537 Q...........FQ.......Q......R........L
33 33 -11.397776 TE....K......V	19 18 -17.716432 K........K......P........H.....T..L
33 33 -11.343313 G..KP...R	19 18 -17.355268 G..K....R...R.....I........R
33 33 -11.295515 A............E.T..........I	19 18 -17.355268 G..K....R..R......I........R
34 33 -11.228934 R...K.....I.......L	19 18 -17.269765 L.FQ.......Q......R........L
33 33 -10.896578 A.R..........L.R	33 18 -17.269765 L...R...........L.FQ.......Q
33 33 -10.819856 K.........AL....R	19 18 -17.200131 R.........R.Q.............LV.............R
33 33 -10.725427 A.R..A...G	19 18 -17.166603 P.....REI....K........A
33 33 -10.725427 A.R..........A....G	35 18 -17.071060 RK....KA..K........K
33 33 -10.691722 E.....L....E.....A	33 18 -17.071060 K....KA.RK........K
33 33 -10.672296 A.K.........GG	18 18 -17.056185 K.T.RK........K.....A
33 33 -10.666422 K....LI.......L	18 18 -16.987304 R...T.......K......V....K.......G
33 33 -10.633829 A....A...G...........T	33 18 -16.886376 K........K......P.........R....T..L
33 33 -10.464820 G.....K.L....A	33 18 -16.794779 K........K......P.T............T..L
49 33 -8.908738 R..........IR	32 18 -16.754280 A..K.........K..R....T.A
33 33 -8.878288 R..RR	19 18 -16.626761 R.Q....AA...A.....V
33 33 -8.832016 K.............RI	32 18 -16.593112 R..G.K..........R.VL
33 33 -8.832016 K....R........I	19 18 -16.570210 T...A.....VT...K....A
35 33 -8.832016 R...K.....I	18 18 -16.461301 G...GKG..K...K
33 33 -8.806310 E..IR	33 18 -16.383679 R..G.K..........R..L..........A
33 33 -8.801566 R..........RK	18 18 -16.073775 L..........A.R........K....L..E
33 33 -8.786690 T.....R........R	20 18 -15.110483 M....Q..R...G..A
48 33 -8.709968 T....RK	30 18 -15.010457 K........K......P.T........Y
33 33 -8.684262 R........E.......T	19 18 -14.872848 L.FQR......Q
33 33 -8.499661 A.R.R	18 18 -14.783276 S...........G..K.HR
33 33 -8.499661 R..........R....A	18 18 -14.746963 A.L............IH..........I
48 33 -8.499661 A.R.............R	20 18 -14.731856 M....Q.A....G..A
34 33 -8.499661 A.R...R	18 18 -14.709361 L......KR....IY
33 33 -8.491037 R............L..Q	31 18 -14.648970 R.Q..T........P....V
33 33 -8.476980 K.............I.G	34 18 -14.585800 P.....REI....K
33 33 -8.446530 R.K.......G	20 18 -14.528393 RFQ..............LV
34 33 -8.446530 K.......G.......R	33 18 -14.489459 K........K......P.T........R
49 33 -8.446530 K...R...G	20 18 -14.479420 I.......I...A.R....R
33 33 -8.446530 K.....G.........R	18 18 -14.409895 R..........RR...........L..Q
33 33 -8.446530 K.............R..G	18 18 -14.363623 R...K.....I.....Q.L
33 33 -8.422939 K......AR	19 18 -14.361940 R.Q..........AP....V
36 33 -8.422939 K.........A.....R	22 18 -14.356331 V.R......E..R.........V
33 33 -8.346226 R......L...I	33 18 -14.331499 E..........E...R..VT
33 33 -8.319885 A...Q.......G	18 18 -14.320889 V..........TV.........K.....I
33 33 -8.315776 R......L.........R	27 18 -14.314817 L.............Y.E......V....V
48 33 -8.238869 T..A.........V	32 18 -14.240702 R..........R.Q..............LV
33 33 -8.213348 E..R......L	19 18 -14.228695 E.....A.....V.....D.....R
36 33 -8.168216 GG.............R	18 18 -14.226020 K.......G..KP...R
33 33 -8.147457 T.....K..........L	19 18 -14.211944 GK...K...KR
33 33 -8.144625 G.A.R	18 18 -14.203938 T.....I.L...IR
34 33 -8.144625 R..........A....G	18 18 -14.181612 AR.....R.S........Q
33 33 -8.091493 G..K.......G	36 18 -14.111631 K....KA..K........K
33 33 -8.067902 A.K..G	36 18 -14.096755 K.T..K........K.....A
50 33 -8.044311 A.....A.K	18 18 -14.085750 A....V.........E.T..........I
33 33 -8.029436 A.T...........A	31 18 -14.081006 K.............T..A.R.......V
33 33 -8.010771 S...K..........A	18 18 -14.071059 L..I.....T..........R...E
48 33 -7.937149 R..L..........A	18 18 -14.060393 G.......R....S....I.........R
33 33 -7.937149 L..........A.R	18 18 -13.983671 G............KS....I.........R
49 33 -7.884018 G.....K.L	34 18 -13.933630 GG.KR.............K
33 33 -7.398229 G...L...........L	19 18 -13.907924 GG.............R...K...E
32 32 -25.862114 L.....D......HA.R.T....D...A....G	30 18 -13.895163 A......T....K......G.......R
32 32 -20.691374 Q..T...IR.......V..I.........R	34 18 -13.891900 Q.........K.P...L....A
32 32 -20.128863 Q..T...I..L.....V..I.........R	18 18 -13.886540 KT.........L..Q.........G
32 32 -19.334493 G.K...K..A.............G..KP	35 18 -13.886448 A..K.........K..R......A
32 32 -17.147792 R..........I.K....RL.R	19 18 -13.886448 RK......A.K.A
32 32 -16.662003 R..........I......RL.R.........L	19 18 -13.856908 G...GK...K...K
32 32 -16.628321 R.G.V........Y......L...L	18 18 -13.856908 GGK...K.......K
32 32 -16.437710 GK...K..A.............G..K	34 18 -13.856908 GKG..K...K
32 32 -14.750874 K..R...........IRR	31 18 -13.841158 V..........K...R.....GG
32 32 -14.448803 R......E...RK......V	19 18 -13.823802 K..A......K.P........A
32 32 -14.172603 A.RK.......V....R	18 18 -13.808926 T...A..........K..A.........P
32 32 -13.253098 A.....A.K..........GG	34 18 -13.802877 R..G.K..........R..L
32 32 -11.474498 TE...R........V	18 18 -13.702564 R....A......K.....L.......K
32 32 -11.436409 G..K....R........I	18 18 -13.687688 T..A.....R....K....L
32 32 -11.333952 T.....R........R.S	20 18 -13.658892 S.E.....L.........I.....T
32 32 -11.267215 G..K........V....I	18 18 -13.610092 L.........A.......K........V.....R
32 32 -11.106886 T....RK..........L	19 18 -13.589995 Q.........AP...L....A
32 32 -11.050923 R.K.......GG	18 18 -13.555003 G..G.K...K.............A
32 32 -11.027332 G..K..........A.R	19 18 -13.547446 K.P...LV...A
32 32 -11.027332 GK.........A.R	31 18 -13.531412 K..........A.....A.....K......G
32 32 -10.950610 G.K...K............A	19 18 -13.488357 A..S...K.P........A
32 32 -10.858138 GV.K..........A	18 18 -13.456033 K..............KP....L.............L
32 32 -10.648705 A.....A.K..G	34 18 -13.447841 K...R.L...GG
32 32 -10.635972 R...K....L.........L	33 18 -13.377002 G..G.G.G.....R
32 32 -8.870353 T.............Q..T	19 18 -13.374953 R....A.....E.....L..........A
32 32 -8.801566 R..........R...K	18 18 -13.332652 TG.....K.L....A
32 32 -8.724844 K......RK	20 18 -13.300280 GGKG....G
32 32 -8.699138 E.R...K	19 18 -13.298230 A...K..........LA.....E
32 32 -8.523252 R..............RG	18 18 -13.296116 G...E.....L....E.....A
32 32 -8.523252 R.....R....G	18 18 -13.178344 G.V.............L.I.......L
32 32 -8.446530 G.............RK	18 18 -13.079904 S.........A..A..V..........A
33 32 -8.446530 GK............R	20 18 -12.941292 K.....L.........T.L.........L
32 32 -8.408063 R...TA	18 18 -12.735726 G.V...........S...L...L

32 32 -8.360293 A.....KP	18 18 -12.591050 H...K.....R..R
32 32 -8.331341 A.K....T	19 18 -12.451682 K.....K..H...P
33 32 -8.144625 G.....A.R	19 18 -12.350046 A.....IH.....I
32 32 -8.067902 K.A.....G	21 18 -12.151054 M....Q.....G..A
32 32 -8.067902 A..G..K	19 18 -12.014325 F.....HA.....L
32 32 -7.991190 L...I.G	20 18 -12.009088 K.....Q....E.....Q
32 32 -7.960740 G.....LR	18 18 -12.002833 R.....LE.....H
32 32 -7.605704 GG.....L	33 18 -11.868167 R...R.....I....R
31 31 -27.811269 K..A.....Q..TK.A.....GGVK	18 18 -11.844658 R.....K.Q..T
31 31 -23.088292 Q..T...IR.L....V..I.....R	31 18 -11.791445 K.....R.IR
31 31 -21.938886 GGK...K..A.....G..KP	34 18 -11.781138 R.Q.....P....V
31 31 -19.544710 R.....I.K....RL.R.....L	34 18 -11.760995 K.....R..R.....R
31 31 -17.515404 R.....IR.....E...R.....V	20 18 -11.719482 L.....KR.....Y
31 31 -16.792746 K...K..A.....G..K....R	34 18 -11.703093 R..P.....EI
31 31 -16.554646 R.....L.....R.V.....T.....A	18 18 -11.698973 V....R.....IR
31 31 -14.433927 R.....TE...R.....V	32 18 -11.689540 Q..T.....P....V
31 31 -14.357205 TE...RK.....V	18 18 -11.669397 T.....RK.....R
31 31 -14.319116 G..K....R.....I....K	33 18 -11.669397 T....RK.....R
31 31 -14.149922 G..K.....V....I....K	18 18 -11.669397 R...T.....RK
31 31 -13.625851 R...K....LI.....L	31 18 -11.641960 P.....Y...T..L
31 31 -13.533380 L.....R....I.....L.....V	34 18 -11.621627 K.....K.....P.....R
31 31 -11.519823 I.....E....K.....V	29 18 -11.591801 R.....K...R.....V
31 31 -11.451155 T.....A...Q..T	18 18 -11.544048 VR.....F.....S
31 31 -11.382368 K.....AR.....R	34 18 -11.530029 K.....K.....P.T
31 31 -11.329237 G.....K..RK	33 18 -11.524582 L.....Y.E.....V
31 31 -11.319722 RK.....PA	32 18 -11.514280 R....T.....P....V
31 31 -11.305656 R.....R.....L...I	35 18 -11.513131 G.....R...R.....I
31 31 -11.121577 T..V.K.....A	36 18 -11.489540 I...A.R....R
31 31 -11.104054 A.R.....RG	35 18 -11.474498 E...R..VT
31 31 -10.615164 S..GK.....A	20 18 -11.436409 G..K....R.....I
31 31 -10.565133 GG.....L.....R	20 18 -11.436409 I.R....G..K
31 31 -9.329716 H..R....A	34 18 -11.405959 R....G..K.....R
31 31 -9.276585 G.....H.K	18 18 -11.405959 K.G....R..R
31 31 -8.775860 R.....E.R	20 18 -11.402511 Q.....AP....V
31 31 -8.699138 R.....E.....K	31 18 -11.397150 V.....QG.T
35 31 -8.632372 R...V.....K	20 18 -11.382368 A.....K.....R..R
31 31 -8.523252 R.....G.....R	19 18 -11.382368 R.....K..R.....A
31 31 -8.523252 G....R...R	31 18 -11.367492 R.....T..A.....R
31 31 -8.523252 G.....R.R	18 18 -11.367492 R.....T.....R.A
34 31 -8.491037 Q.....R.L	33 18 -11.358869 R....T..L.....Q
31 31 -8.446530 G.....K..R	34 18 -11.348038 E.....L.R.....Q
48 31 -8.431654 T.....R.G	32 18 -11.343937 G.V.....R.....I
31 31 -8.389398 S...K.....R	18 18 -11.333952 T.....R....S.....R
31 31 -8.346226 R.....I..L	32 18 -11.313487 G.....R...V.R
31 31 -8.346226 L...IR	18 18 -11.305656 R.....I.....RL
31 31 -8.213348 R.....E.....L	19 18 -11.305646 K.....A.R...K
31 31 -8.168216 G.....G....R	31 18 -11.290770 A.....T...K.....R
31 31 -8.053027 T.....A....G	20 18 -11.290770 T..K.....A.R
31 31 -7.845552 L.....A...T	24 18 -11.289896 ARR..V
31 31 -7.756048 S....GG	34 18 -11.281272 R.Q.....LV
30 30 -23.477955 V.....T.....K.V..M.....L.R..R	31 18 -11.279314 Q..R...G..A
30 30 -22.293923 G.K...K..A.....G..KP...R	18 18 -11.272106 K...RKS
30 30 -20.398111 R.....IR.....E...RK.....V	18 18 -11.257230 K.....T.....R.....S
30 30 -19.397139 GK...K..A.....G..K....R	19 18 -11.257230 KST.....R
30 30 -19.203740 G.K..RK.L.....PA	31 18 -11.255723 Q.AR.....A
30 30 -17.316634 R.....TE...RK.....V	19 18 -11.252515 G.K...K...K
30 30 -16.685399 R.....T.GV.K.....A	18 18 -11.252515 GK...K....K
30 30 -16.492809 L.R.....R....I.....L.....V	18 18 -11.252515 K...K.....K.....G
30 30 -15.359038 D.....H.K.....D.....G	20 18 -11.252515 G....K...K...K
30 30 -14.226645 G..K....R...V....I	18 18 -11.252515 GK...K.....K
30 30 -13.593258 A.R..A...G.....T	35 18 -11.252515 GK...K...K
30 30 -12.266464 G.....H.K.....I	19 18 -11.252515 G.....K...K.....K
30 30 -11.658567 R.....E.R...K	18 18 -11.252515 G.K...K.....K
30 30 -11.504948 I.....TE.....V	19 18 -11.243000 KA.....K.....P
30 30 -11.436409 K.....RI.G	18 18 -11.237639 T.....GK...K
30 30 -11.405959 G.....RK.....R	18 18 -11.228924 K...K..A....K
30 30 -11.127645 G.....G....R...R	20 18 -11.214048 A...T..K...K
30 30 -10.970201 S...K.....A.R	32 18 -11.198484 R.....KR....L
30 30 -10.372349 A...G.....L.....V	18 18 -11.198299 A.....V.....T.....R
31 30 -9.673306 K....M.....R	18 18 -11.198299 T..AR.....V
30 30 -9.631621 H.KR	33 18 -11.198299 T..A.R.....V
30 30 -8.806310 R.....I..E	18 18 -11.158095 G.....G.....I.R
30 30 -8.775860 R.....R.....E	18 18 -11.158095 G.....G....R.I
30 30 -8.523252 R.....G.....R	19 18 -11.158095 G.....G....R.....I
30 30 -8.431654 R.....T.G	18 18 -11.130292 T.....VT.....G
30 30 -8.330467 V.....A.R	18 18 -11.096056 K.....E..R.....L
30 30 -8.254629 T.....L...I	20 18 -11.086530 Q.A.....A.....V
30 30 -8.144625 A.R.....G	20 18 -11.070514 R...R.G.....S
30 30 -8.067902 A....K.....G	19 18 -11.050923 KG....G.....R
30 30 -8.044311 K.....A.....A	20 18 -11.050923 GG.....R...K
30 30 -8.029436 A...T.....A	35 18 -11.050923 K...R.....GG
30 30 -7.951840 AV.....A	18 18 -11.045039 K..K.....K.....L
31 30 -7.860427 K.....L.....A	18 18 -11.043980 V.....V.K.....A
30 30 -7.767955 V.AL	31 18 -11.027332 K.....R...G..A
29 29 -26.345787 V.....T.....K.VT.M.....L.R..R	35 18 -11.009193 Q.....P...L....A
29 29 -24.898316 GGK...K..A.....G..KP...R	33 18 -11.003741 A.....RK.....A
29 29 -20.715385 L.....D.....H.KR.....D.....G	18 18 -11.003741 AA.....K.....R
29 29 -20.383236 R.....IR.....TE...R.....V	18 18 -11.003741 A..A.....K.....R
29 29 -18.226870 D.....H.K..T....D.....G	20 18 -10.993792 G.KR.S
29 29 -18.185323 G.....H.K.....I.....R..R	22 18 -10.988901 G.....GV..I
29 29 -17.939840 D.....HAK.....D.....G	20 18 -10.974201 G...GK...K
29 29 -17.109352 G..K....R...V....I....K	19 18 -10.974201 G..G.K...K
29 29 -16.869478 GL.....R..I.....RR	34 18 -10.974201 G....GK...K
29 29 -14.387655 I.....TE....K.....V	19 18 -10.969411 Q....A.L....E
29 29 -14.288666 G.....K..RK.....R	19 18 -10.950610 A.K.....G..K
29 29 -13.574594 S..GK.....A.R	20 18 -10.950610 K...K.A....G
29 29 -11.482681 R.....G.....R.R	35 18 -10.941095 A.....K.P.....A
29 29 -11.289896 V.....A.R...R	31 18 -10.934860 A.R..V.....G
29 29 -11.214058 T.....L...IR	35 18 -10.927019 K.....A.K.A
29 29 -11.172778 R.....R.....E.....L	19 18 -10.927019 K.....KAA
29 29 -11.027332 A...K.....G.....R	20 18 -10.927019 A...A.....KK
29 29 -10.834547 A..K...A...V	18 18 -10.912143 A..A...T...K
29 29 -10.702553 E.....L.....A...T	18 18 -10.911269 A....RV.....A
29 29 -8.739544 V.....IR	18 18 -10.911269 A..A.RV
29 29 -8.709968 K.....R.....T	19 18 -10.902645 Q.....LV.....A
29 29 -8.044155 E.....V.L	18 18 -10.893478 K...K..A.....S
29 29 -7.860427 K.....AL	34 18 -10.893478 A.KS.....K

28 28 -23.583217 L.....D......H.KR.T....D........G	18 18 -10.867147 Q.........A..G.....S
28 28 -23.296187 L.....D......HAKR......D........G	18 18 -10.843447 G........K.....LR
28 28 -23.265943 R..........IR.....TE...RK......V	18 18 -10.833932 R..P...L....A
28 28 -20.807672 D......HAK..T....D........G	18 18 -10.819856 L....A..........K.....R
28 28 -20.807672 D......H.K..T....D...A....G	20 18 -10.808841 E.............A..A..V
28 28 -13.793976 A..K...A...V....R	21 18 -10.796357 L..............T.L..........Q
28 28 -8.632372 K........V.....R	34 18 -10.772609 G.G.G.....R
29 28 -8.422939 KR..........A	33 18 -10.772609 GR...G.G
28 28 -8.331341 A.T.....K	18 18 -10.766725 K....G..K.....L
28 28 -8.253745 A.......K........V	18 18 -10.743144 L.IR......L
28 28 -8.029436 T..A.....A	20 18 -10.743134 A.......K....K.....L
28 28 -7.975431 G........A.....V	18 18 -10.742075 A....V.A..V
28 28 -7.791546 V..........L.G	18 18 -10.718761 Q....L.........LV
28 28 -7.791546 G..........V.L	18 18 -10.717428 L....E.....A.........K
27 27 -26.164019 L.....D......HAKR.T....D........G	20 18 -10.702553 A....L.E............T
27 27 -26.164019 L.....D......H.KR.T....D...A....G	33 18 -10.695887 G.GK.G
27 27 -23.388474 D......HAK..T....D...A....G	33 18 -10.695887 GKGG
27 27 -16.753405 A.RK...A...V....R	35 18 -10.695887 GKG....G
27 27 -13.794850 A.T.....K.....A.......K	19 18 -10.693844 V.R.S.............L
27 27 -13.555003 K..........A.K.........GG	33 18 -10.693844 R...S...........LV
27 27 -11.213174 A.......K........V.....R	19 18 -10.693844 S............LVR
45 27 -10.927019 A.......K.....A.......K	19 18 -10.693212 Q.A.......A....L
27 27 -10.750976 G..........V.L....R	18 18 -10.681011 T..G.GG
27 27 -8.709968 R.........K....T	18 18 -10.672296 GGK.........A
27 27 -8.607540 E.......T.....K	19 18 -10.659672 Q....S....A....L
28 27 -8.422939 RK...A	19 18 -10.650663 K.........VA............L
27 27 -8.422939 R.........K.....A	19 18 -10.648705 A...A....G..K
27 27 -8.360917 I.............V.A	35 18 -10.648705 A....G.K..........A
27 27 -7.975431 V.A..........G	18 18 -10.648705 A.......K....A....G
26 26 -28.744821 L.....D......HAKR.T....D...A....G	19 18 -10.624957 VAL.E
26 26 -21.584439 A.T.....K.....A.....A.K.........GG	22 18 -10.591417 S..............V.L.E
26 26 -11.314361 R.........K....T...........G	32 18 -10.565133 L.R.G......G
26 26 -8.808889 M....A.......L	18 18 -10.543500 R.....L.........LV
26 26 -8.446530 K.............R.G	20 18 -10.499102 S...........A..V.........A
26 26 -8.331341 A...T...K	19 18 -10.499102 S........A.....V.........A
26 26 -8.069860 K.........V..L	19 18 -10.464820 A........G..K...........L
25 25 -14.113066 R.....A.D....L..Q	18 18 -10.441229 K..........LA...A
25 25 -13.585260 E.....L..........A...T...K	33 18 -10.417573 G....G.GG
25 25 -11.669397 R.........K....T.........R	33 18 -10.417573 G.G.G..G
25 25 -11.577588 F...V...A.......R	20 18 -10.415864 L.............D..LA
25 25 -11.566969 R.......E.......T.....K	19 18 -10.407689 A.......KS.........L
25 25 -10.965310 I.............V.A..........G	20 18 -10.242469 L.............T.L....A
25 25 -8.540774 T...K.........V	19 18 -10.242469 L.........A....T.L
40 25 -8.491037 R...........L..Q	21 18 -10.162915 A.....G......L....A
27 25 -8.397574 R.......D....L	18 18 -9.955440 A.......A.......L.....L
24 24 -14.813726 S.....G..........H.K........I	24 18 -9.399286 R..........Y...R
24 24 -14.273791 R.........K....T.........R.G	35 18 -9.322564 KR.....Y
24 24 -10.650663 K.........V.AL	19 18 -9.321092 H........L......Q
25 24 -8.253745 K.........V.A	18 18 -9.296176 S..........H..R
24 24 -8.144625 R...........A....G	19 18 -9.276585 K.......G....H
23 23 -20.732585 S.....G..........H.K........I......R..R	18 18 -9.276585 K.......G....H
23 23 -14.434589 E.......F...V...A.......R	18 18 -9.145832 H..R.....L
23 23 -14.382911 K.......R.....T...K.........V	19 18 -9.119708 I.K.............F
23 23 -11.436409 I....K...R......G	18 18 -9.083998 I.....Q.........R
23 23 -11.382368 AR.K.............R	34 18 -9.083998 I.....Q...R
23 23 -8.739544 R.V..I	27 18 -9.060898 Y........V....V
23 23 -8.637116 E......I....V	22 18 -9.016364 M...........G..A
23 23 -8.633918 I..L...........F	18 18 -8.990535 R..........I.D
23 23 -8.606666 V...E..R	22 18 -8.981571 I...Q...E
40 23 -8.446530 K...R......G	18 18 -8.976826 K.......R...Q
23 23 -8.408063 AR..T	31 18 -8.939188 I..........IR
23 23 -8.346226 L.I.......R	18 18 -8.908738 R.......I.......R
23 23 -7.967599 AI..L	35 18 -8.908738 R......I........R
22 22 -14.916836 I....K...R......G........Y	22 18 -8.908738 IR.........R
22 22 -11.596546 E......I....V...........R	19 18 -8.908738 I........R..R
22 22 -11.474498 V...E..R....T	21 18 -8.885229 T.......K....Q
22 22 -11.407732 V......R..TV	20 18 -8.883363 R...K........D
22 22 -7.951840 A....A....V	18 18 -8.878288 R.............R..R
21 21 -16.716034 K.....L...I....K...R......G	18 18 -8.832016 K..............I.R
21 21 -15.904748 L....A.......LR.....I........A	18 18 -8.832016 R........I.K
21 21 -14.290439 V.......R..TV.........K	36 18 -8.821709 Q..........P....V
21 21 -13.880288 T.....R....T.......A....G	19 18 -8.801566 K..............RR
21 21 -13.784371 L........R....I..D....L	48 18 -8.801566 RK.......R
21 21 -13.507664 K..L...........E.A....V	19 18 -8.801566 K.............R.R
21 21 -11.593347 R.....I..L...........F	32 18 -8.790891 V.RD
21 21 -11.310390 E.....AI.........K	18 18 -8.786690 R..........R...T
21 21 -11.305656 I.........RRL	18 18 -8.786690 R...T..........R
21 21 -11.271316 Q..K..L...........E	34 18 -8.786690 T....R.........R
21 21 -10.514861 AI...L.........S	18 18 -8.780310 G..........D..Q
21 21 -9.341240 R........F......Q	19 18 -8.775860 E.........R..R
21 21 -9.053548 RR.............Q	18 18 -8.769994 V..I...I
21 21 -8.611908 I..D.............A	19 18 -8.740418 T.......K........I
21 21 -8.521487 I.....Q.L	18 18 -8.740418 T..........K.....I
21 21 -8.437472 E.V....V	34 18 -8.739544 V..............IR
21 21 -8.422939 A.R........K	18 18 -8.738920 K...R..P
21 21 -8.422939 A.....R....K	49 18 -8.738920 K.......P.........R
23 21 -8.408063 R....T.......A	18 18 -8.724844 R....K...K
21 21 -8.397233 R.......A..E	34 18 -8.724844 K...K..R
22 21 -8.389398 KS.......R	51 18 -8.724844 RK.........K
21 21 -8.374523 T.........R.....S	37 18 -8.724844 K.........K..R
21 21 -8.346226 I.L....R	18 18 -8.724844 K...K............R
21 21 -8.330467 V...........R....A	33 18 -8.709968 K.........T.......R
21 21 -8.330467 AR........V	24 18 -8.709094 V.R.........R
21 21 -8.315776 L....R.........R	34 18 -8.709094 V.............R..R
21 21 -8.294805 E...A......E	32 18 -8.699294 T.............D.V
21 21 -8.254629 T.....I.L	32 18 -8.698512 Q..R...G
21 21 -8.224841 A...L...........F	18 18 -8.688463 D.V...E
21 21 -8.121033 A.R.......A	32 18 -8.674921 Q.AR
20 20 -23.079169 K.......K.....L...I....K...R......G........Y	36 18 -8.662198 K.........K......P
20 20 -19.571809 E.V....V...E...........V....K.........G	36 18 -8.648122 K...K..K
20 20 -18.892500 S...........K.A.......GGVK	19 18 -8.647947 TV....I
20 20 -18.864178 R......L....A......LR.....I........A	18 18 -8.637116 V.............E..I
20 20 -18.692859 A..K.....G.V...........A..E.........L	33 18 -8.633246 TK.............K
20 20 -17.464098 I....K...R......G....S...Y	19 18 -8.632372 R...K.....V
20 20 -14.426288 I.K..I....R..G	37 18 -8.612275 Q.........AP
20 20 -14.300269 E.....AI.........K.I	33 18 -8.606915 QG.T
20 20 -14.188363 K.....I......RL.R	33 18 -8.606666 V.R......E

20 20 -14.154023 K..........Q..K..L..........E	18 18 -8.581458 D...A.....R
20 20 -13.581188 R.......AL.........LF	33 18 -8.553702 I.G......R
20 20 -13.229507 KA........A....A...G	18 18 -8.530111 A......R.........I
20 20 -11.571337 R.I..D.............A	22 18 -8.530111 I......R.A
20 20 -11.450466 RR..........L..Q	36 18 -8.523252 R...R.G
20 20 -11.404194 K.....I.....Q.L	34 18 -8.523252 G.........R.....R
20 20 -11.382368 K.A.R....R	19 18 -8.499661 R..A......R
20 20 -11.290770 T..A.....R....K	33 18 -8.499661 R........R..A
20 20 -11.289896 VR.........R....A	19 18 -8.499661 A....R..........R
20 20 -11.214058 T.....I.L....R	18 18 -8.499661 R...........A.R
20 20 -11.198484 K....L.......R..R	34 18 -8.499661 A..R..R
20 20 -11.086530 Q.....A....A....V	35 18 -8.491037 L.R..........Q
20 20 -10.927029 R..............AI..L	21 18 -8.491037 R...Q.....L
20 20 -10.875607 E...A......E.....A	18 18 -8.466121 R.........R..S
20 20 -10.743134 A......K.....L......K	18 18 -8.446530 G.......K.......R
20 20 -10.691722 E...A.L....E	18 18 -8.446530 K.....G......R
20 20 -9.662071 K.H..........I	32 18 -8.446530 RG..K
20 20 -9.094002 F....I.......E	36 18 -8.446530 G.KR
20 20 -9.053548 Q.......RR	18 18 -8.446530 R.K.G
20 20 -8.929061 A..Y.......T	33 18 -8.446530 G..K.......R
21 20 -8.878288 R.........RR	19 18 -8.446530 G..............KR
33 20 -8.832016 K..............IR	32 18 -8.446530 K......G.......R
20 20 -8.724844 K..K........R	34 18 -8.446530 G..K.............R
20 20 -8.724844 K..........K....R	34 18 -8.446530 K...........R...G
20 20 -8.709968 K......T.....R	34 18 -8.446530 R....G..K
20 20 -8.709094 V..........RR	18 18 -8.446530 K..GR
21 20 -8.504238 E.V........E	18 18 -8.446530 G.K...R
20 20 -8.496571 S....I.........R	18 18 -8.422939 K...........R...A
20 20 -8.491037 Q.........RL	32 18 -8.422939 K.......R..A
20 20 -8.422939 K...R....A	21 18 -8.422939 K...R...........A
20 20 -8.389398 KS..............R	35 18 -8.422939 K.....A.R
20 20 -8.376677 I.L...I	18 18 -8.422939 K.........A....R
20 20 -8.363693 S.E...........R	36 18 -8.422939 K......A.R
20 20 -8.327377 S..I......V	18 18 -8.422939 K.......R.A
20 20 -8.318396 G...E..........E	22 18 -8.422939 R.K.........A
20 20 -8.312676 K.....S.......K	18 18 -8.422939 A.R..............K
22 20 -8.239054 K....L.......R	18 18 -8.422939 RK.........A
20 20 -8.176408 KP....L	34 18 -8.422939 K..........A.R
20 20 -8.175075 I.........A...G	19 18 -8.422939 A.....R...K
20 20 -8.161273 A....V....V	33 18 -8.408063 T..A...........R
20 20 -8.121033 A.............A.R	18 18 -8.397233 E...R.A
20 20 -8.067902 A...G..K	19 18 -8.389398 R............KS
20 20 -8.044155 V...E.........L	32 18 -8.389398 RKS
20 20 -8.044155 LV.E	18 18 -8.389398 R.S...K
20 20 -7.977389 V........L.......V	32 18 -8.389398 R.S.......K
24 20 -7.967599 I.A...........L	35 18 -8.369808 G.....K...K
20 20 -7.903609 S......R..L	35 18 -8.369808 G.K.............K
21 20 -7.834721 A.L....E	18 18 -8.369808 GK.....K
20 20 -7.801181 S.............L..E	20 18 -8.369808 G.K.......K
20 20 -7.789588 A...GG	34 18 -8.369808 G.....K.......K
20 20 -7.732457 A..G.....S	37 18 -8.369808 G....K...K
20 20 -7.558522 L........A.....A	34 18 -8.369808 K.......G..K
21 20 -7.374638 AL..........L	18 18 -8.369808 G.........K...K
20 20 -7.341097 S...L...L	36 18 -8.360293 K..A........P
19 19 -142.36677 STELLI...PFQRLV.EIAQDFKT.LRFQ..A..ALQE..EA..V.LFEDTNL.AIH.K.V....KD..L. ....GER	37 18 -8.360293 A.......K......P
19 19 -27.733673 K..A...........Q...K.A.......GGVK.........V	18 18 -8.346226 I.......L.........R
19 19 -22.350520 T...IR.L.....V..I.........R......L	19 18 -8.346226 R........I.......L
19 19 -21.851930 R.S.............K.A.......GGVK	33 18 -8.346226 L.R.I
19 19 -20.516114 T...IRR......V..I.........R	21 18 -8.346217 A..........KK
19 19 -19.796693 L.............R..IQ.......RRL	36 18 -8.346217 A.K........K
19 19 -17.638482 T...I.........V..I......D.R	36 18 -8.346217 K........K.A
19 19 -17.071070 K.....I.K....RL.R	36 18 -8.346217 A..K.........K
19 19 -16.769165 A........K.....I......RL.R	18 18 -8.331341 A..K.............T
19 19 -16.743800 L.R.....R....I..D....L	19 18 -8.331341 A....T.......K
20 19 -16.501018 Q.............L..Q..A...L....E	33 18 -8.331341 K.......T.A
20 19 -16.405331 P......EI.....T.L.........L	33 18 -8.331341 T......A.......K
19 19 -16.364665 E.....K..L...........E.A....V	35 18 -8.331341 A......T...K
19 19 -14.963269 R.........F..D...Q.............A	31 18 -8.331341 A......T....K
20 19 -14.872848 FQ.......Q......R........L	18 18 -8.330467 R..V.............A
34 19 -14.872848 R.............L.FQ.......Q	35 18 -8.330467 A.R..V
34 19 -14.246320 Q...E.....L........I.....T	35 18 -8.330467 A.R........V
20 19 -14.207173 P......EI....K........A	18 18 -8.330467 R..V.....A
19 19 -14.188363 K....L.......R..R.I	18 18 -8.330467 R.AV
19 19 -14.172603 K......VR..........R....A	36 18 -8.321843 Q..............LV
19 19 -14.124383 A....I....I....R..G	19 18 -8.319885 Q...........A.G
35 19 -14.045959 R.Q.....A...A.....V	18 18 -8.315776 L....R..........R
19 19 -14.017211 A...G..K...R..........I	22 18 -8.312676 K...K.........S
19 19 -13.924739 I.........A.R.G..............V	18 18 -8.286970 E.....KS
32 19 -13.575919 K.........QLA........A	18 18 -8.262460 V.........T....G
19 19 -13.474291 R..............AI..L.........S	19 18 -8.254629 L..I.....T
19 19 -13.323936 A.......K.....LA.....K	24 18 -8.239054 R.............LK
19 19 -13.272525 E...A.L....E.....A	35 18 -8.239054 K...L..R
19 19 -13.214631 A....A...G...........T.A	18 18 -8.239054 KL.......R
23 19 -13.129203 A............AI.A...........L	19 18 -8.239054 L...R.........K
34 19 -12.996325 A.....EA....L..........A	19 18 -8.238869 A.....V.........T
32 19 -12.812441 A...............AL.E.............L	18 18 -8.228380 D.V..L
19 19 -12.710177 A....A...G....S..........L	34 18 -8.224179 R....T..L
19 19 -12.591050 H.KR..............R	18 18 -8.213348 R.......E.........L
35 19 -12.475930 R..............FQ.......Q	21 18 -8.177033 V............L..I
20 19 -12.475930 FQR......Q	19 18 -8.169291 D........S....A
21 19 -12.131476 RFQ..............V	34 18 -8.168216 G......G.....R
21 19 -12.048687 R.........K...RF	19 18 -8.168216 G...G.........R
19 19 -12.012978 R.........RR..............Q	18 18 -8.168216 G..G.............R
37 19 -11.913418 L.FQ.......Q	34 18 -8.162332 K.........K.....L
19 19 -11.760995 R.........R...K........R	18 18 -8.151484 A.........I..A
19 19 -11.757084 A.L...........IH	18 18 -8.150702 P.........L.E
20 19 -11.678343 E...........V..I..D	18 18 -8.147457 L......K....T
21 19 -11.673599 V.....D.K...R	29 18 -8.144625 G........A....R
19 19 -11.668523 RR..............R.V	47 18 -8.144625 R...G..A
21 19 -11.646560 T..L.....FQ	18 18 -8.121033 A...........A.R
36 19 -11.626371 P......EI....K	33 18 -8.121033 AR...........A
19 19 -11.621627 K.......RK..........P	21 18 -8.121033 R....A.......A
20 19 -11.573200 L..IH...........L	51 18 -8.091493 GKG
37 19 -11.465157 R.Q.....A.........V	36 18 -8.091493 KG....G
36 19 -11.465157 R.Q.........A.....V	19 18 -8.091493 G..G......K

19 19 -11.379278 KS....I.........R	34 18 -8.091493 KGG
35 19 -11.378488 Q...E.....L.........I	34 18 -8.091493 G..K.G
19 19 -11.336106 I.L...IR	35 18 -8.091493 GK.G
19 19 -11.289896 V.........ARR	20 18 -8.091493 G..G.K
19 19 -11.272106 KS......R.........K	18 18 -8.087493 A.R.......S
20 19 -11.269266 E.....A.....V.....D	36 18 -8.067902 A........G..K
33 19 -11.265064 T..........R...E.A	20 18 -8.067902 A...G.K
20 19 -11.252515 GK...K........K	19 18 -8.067902 K....A....G
19 19 -11.244508 T.....I.L...I	32 18 -8.044311 K.........A.....A
19 19 -11.222183 A......R.S........Q	21 18 -8.044311 A...A........K
19 19 -11.222183 AR.......S........Q	20 18 -8.044311 AA..........K
34 19 -11.164126 T............Q.A..A	20 18 -8.044311 A.........K.....A
33 19 -11.135837 K.....P....L....R	20 18 -8.044311 A..K.........A
19 19 -11.134504 A..G....I.........R	18 18 -8.034362 K.....S..G
19 19 -11.132720 GG.K.......D	21 18 -8.019487 STG
20 19 -11.121577 T...A.....V....K	18 18 -8.010771 K.........A...S
36 19 -11.111630 E.....L.........I.....T	18 18 -8.010771 A......S........K
22 19 -11.081180 T..........RL..E	18 18 -8.010771 A..K.........S
35 19 -11.027332 A.R.....K.........G	19 18 -8.010771 K......A......S
19 19 -11.027332 G.........KA..............R	32 18 -7.999022 G.............GV
19 19 -11.014414 R......L.......VT	19 18 -7.995896 A........ST
34 19 -10.995117 K.........QL.........A	21 18 -7.995896 A........ST
33 19 -10.995117 K.........QLA	18 18 -7.975431 V.......A....G
23 19 -10.969411 Q..A...L....E	35 18 -7.960740 L.R.G
19 19 -10.927029 AI..L.......R	19 18 -7.937149 A.R..L
19 19 -10.927029 I......A...L..R	21 18 -7.928525 L.........Q.....L
19 19 -10.911986 V...E.......T.L	34 18 -7.884018 G..K.....L
32 19 -10.900687 Q.A....G..A	20 18 -7.860427 L....A..........K
19 19 -10.899198 G...E..........E.....A	20 18 -7.860427 A..........K....L
31 19 -10.897268 T...A......T.........A	19 18 -7.860427 A............L..K
19 19 -10.878603 S...........T..A..........K	21 18 -7.845552 A..........T.L
19 19 -10.863038 R.......S.......R..L	20 18 -7.834721 E.....L.........A
19 19 -10.845220 V.......L.......VT	49 18 -7.813179 G.G.G
20 19 -10.805643 L....A.........F......A	18 18 -7.813179 G...G......G
20 19 -10.785527 L..........QE........L	50 18 -7.813179 G.G..G
36 19 -10.757210 K.P...L....A	35 18 -7.813179 G..G.G
19 19 -10.728259 T......K.L....A	66 18 -7.813179 G....G.G
19 19 -10.715313 G...E.....L....E	34 18 -7.813179 G...G..G
20 19 -10.702553 AL.E............T	35 18 -7.813179 G....G..G
20 19 -10.664739 P...LV...A	50 18 -7.813179 G.GG
19 19 -10.650663 L.........A.......K........V	50 18 -7.813179 G.G....G
35 19 -10.599408 A.....EA..............A	19 18 -7.767955 L......AV
24 19 -10.548401 A.............AI.............L	30 18 -7.765997 G.........A..A
34 19 -10.532642 A..A..V.........A	21 18 -7.753265 L...L...R
21 19 -10.518669 TEL.........L	35 18 -7.753265 L......R.........L
33 19 -10.415523 A...............AL.E	19 18 -7.753265 R......L.........L
20 19 -10.415523 A............A...L.E	21 18 -7.742406 A..A....A
35 19 -10.415523 A.....EA....L	19 18 -7.742406 A.........A...A
36 19 -10.415523 A.....E.....L..........A	19 18 -7.734415 S.L.......V
19 19 -10.364517 L.........I.A...........L	18 18 -7.708866 S..........A...A
19 19 -10.338808 G.V..........S...L	18 18 -7.676543 K...L....L
35 19 -10.231639 AL.E.............L	32 18 -7.605704 G..L...G
23 19 -10.231639 L............E.A......L	20 18 -7.558522 LA...A
20 19 -10.231639 AL.E........L	20 18 -7.558522 A.......A....L
19 19 -10.139324 A....L........A..A	20 18 -7.558522 L....A...A
37 19 -10.139324 A.............A..A...........L	19 18 -7.374638 L.....A.L
36 19 -10.058585 L.........T.L.........L	18 18 -7.374638 AL............L

REFERENCES

[1] Rigoutsos, I. and A. Floratos, "*Teiresias*: An Algorithm For The Discovery Of Patterns In 1-Dimensional Event Streams." IBM Thomas J. Watson Research Center Technical Report. Spring 1997.

[2] Reeck GR, E. Swanson, and D.C. Teller, "The Evolution Of Histones," Journal Of Molecular Evolution, 10, 309-317, 1983.

[3] Ouzounis C.A., and N.C. Kyrpides, "Parallel Origins Of The Nucleosome Core And Eukaryotic Transcription From Archaea," Journal Of Molecular Evolution, 42, 234-239, 1996.

[4] Bairoch A, P. Bucher, and K. Hofmann, "The PROSITE database: its status in 1995," Nucleic Acids Research, 24, 189-196, 1996.

[5] Rigoutsos I., A. Floratos and C. Ouzounis, " Case Studies in Pattern Discovery Wihout Alignment: Results Using the Teiresias Algorithm." IBM Thomas J. Watson Research Center Technical Report. Spring 1997.

[6] Sidman K.E., D.G. George, W.C. Barker, and L.T. Hunt, "The Protein Identification Resource (PIR)," Nucleic Acids Research, 16(5):1869-71, 1988.