

## SEARCHING FOR PALINDROMIC SEQUENCES IN PRIMARY STRUCTURE OF PROTEINS

MARCIN HOFFMANN<sup>1</sup> AND JACEK RYCHLEWSKI<sup>1,2</sup>

<sup>1</sup>*Quantum Chemistry Group, Department of Chemistry, A. Mickiewicz University,  
ul. Grunwaldzka 6, 60-780 Poznań, Poland*

<sup>2</sup>*Poznań Networking and Supercomputing Center, Institute of Bioorganic Chemistry,  
ul. Wieniawskiego 17, 61-713 Poznań, Poland*

**Abstract:** Protein data base SWISSPROT was tested in the search for palindrome sequences in primary structure of polypeptides. The obtained results indicate that palindromic words are present in protein structure and there is a number of them. Half of the length of the longest palindrome was 76 and in accordance with expectations the shorter the length of the palindrome the greater number of them has been determined.

### 1. INTRODUCTION

Palindrome is a word or phrase which reads the same backwards as forwards, for example *radar* [1], Palindromic sequences among nucleic acids are well known and of utmost importance [2-6]. Restriction endonucleases, discovered in late sixties, usually recognize palindromic sequences of DNA and after the recognition they cleave the DNA [1]. Restriction enzymes which recognize specific sequences of DNA are widely exploited in laboratories for gene isolation and cloning, genetic recombination, examining of chromosome structure, and sequencing of long DNA fragments [1,7-10]. Some nuclear receptors recognize as their target palindromic sequences of DNA [11-16]. For example the nuclear glucocorticoid receptor recognizes a palindromic fragment of human DNA [17,18]. Moreover palindromic sequences are present in telomeres and are necessary for initiation of DNA replication therein [19,20],

All in all palindromic sequences of DNA were studied in depth and their role is quite well understood [1-20]. On the other hand the presence of palindromic sequences in primary structure of proteins has not gained as much attention [21,22]. Therefore an attempt to search for palindromic sequences in polypeptides have been undertaken.

### 2. METHODS

Protein database, SWISSPROT [23], was chosen to be examined during the search for palindromic sequences. A computer program which was searching for palindromic words was written based on an algorithm:

1. read in the sequence of the protein and the length of the polypeptide chain,
2. choose the initial values of the length of a half of palindrome LH (largest value, equaled 99) and the length of a gap in the middle of palindrome LG (initially 0),
3. start from the first aminoacid,
4. check whether a sequence of the length of  $2 * LH + LG$  is a palindrome with the gap in the middle of the length LG

- if yes write the name of the protein and the position of the beginning and the end of a palindrome, move to next step (number 5),
  - if no move to next step (number 5),
5. move the starting position of one aminoacid and repeat 4 until the end of polypeptide chain is reached,
  6. increment by one LG and go to step 3 (until LG is lower than 6),
  7. decrement by one LH and go to step 3 (until LH is greater than 2),
  8. after checking one protein sequence check the next protein.

### 3. RESULTS

The number of observed palindromes in sequence of proteins and half of the length of these palindromes are presented in Table I. The Chart 1 presents the number of palindromes in logarithmic scale versus their half-lengths.

Table I. Number of observed palindromes and half of their length

Half-length of palindrome	No. of observed palindromes		Half-length of palindrome	No. of observed palindromes		Half-length of palindrome	No. of observed palindromes	
	total	exclusive		total	exclusive		total	exclusive
(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
77	0	0	52	113	5	27	1047	94
76	1	1	51	119	6	26	1148	101
75	3	2	50	126	7	25	1255	107
74	7	4	49	134	8	24	1374	119
73	13	6	48	143	9	23	1507	133
72	18	5	47	152	9	22	1661	154
71	23	5	46	161	9	21	1841	180
70	28	5	45	172	11	20	2043	202
69	32	4	44	185	13	19	2268	225
68	36	4	43	201	16	18	2526	258
67	40	4	42	222	21	17	2817	291
66	44	4	41	250	28	16	3147	330
65	48	4	40	281	31	15	3549	402
64	52	4	39	315	34	14	4028	479
63	56	4	38	352	37	13	4632	604
62	62	6	37	392	40	12	5371	739
61	70	8	36	433	41	11	6323	952
60	76	6	35	479	46	10	7560	1237
59	82	6	34	528	49	9	9300	1740
58	86	4	33	586	58	8	11839	2539
57	92	6	32	648	62	7	15882	4043
56	96	4	31	717	69	6	22430	6548
55	100	4	30	792	75	5	35235	12805
54	104	4	29	868	76	4	65853	30618
53	108	4	28	953	85	3	202558	136705

Number of observed palindromes exclusive - the palindromes of the greater length are not counted

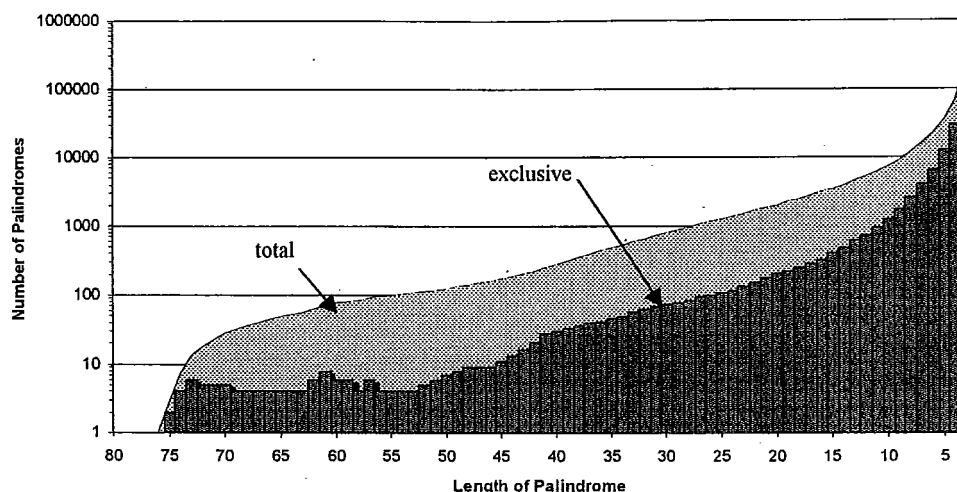


Chart 1. Number of observed palindromes and half of their length

As can be easily noticed the longest observed palindrome has the length of  $2 \cdot 76 = 152$  aminoacids, whereas as expected the number of observed palindromes increases as the length of palindrome decreases. Number of palindromes of the length  $2 \cdot 3 = 6$  is 202 558 if counting all hits and 136705 if longer palindromes are excluded.

#### 4. DISCUSSION AND CONCLUSIONS

The obtained results indicate that in primary structure of polypeptides one can observe palindromic sequences of various length. As could be expected the shorter the palindrome the more frequently it occur. The longest palindrome sequences are monotonous - they are build from one or two aminoacids. The role of palindromes for structure or function of the protein remains still unclear and definitely deserves further examination. Whether their presence is caused by accident or these palindromic sequences are underlying some structural or functional features is currently under investigation and further results are to be published soon. All in all palindromic words of aminoacids do exist in the sequence of proteins and examining of their role would be an exciting challenge.

#### Acknowledgement

Authors thank Prof. J. Barciszewski for fruitful discussions and for stimulating interest in finding palindromic sequences in protein primary structure. The support from Poznań Networking and Supercomputing Center is gratefully acknowledged.

#### References

- [1] L. Styer, *Biochemistry*, 4th Ed., W. H. Freeman & Co: New York (1995).
- [2] C. Chauveau, C. Decourt, M. Cogne, *Gene* **222**, 279 (1998).
- [3] M. Gupta, R. Zak, T. A. Libermann, M. P. Gupta, *Molec. Cell. Biol.* **18**, 7243 (1998).

- [4] S. Shankar, B. Mohapatra, A. Suri, *Biochem. Biophys. Res. Comm.* **243**, 561 (1998).
- [5] D. G. Taghian, H. Hough, J. A. Nickoloff, *Genetics* **148**, 1257 (1998).
- [6] S. R. Bhaumik, K. V. R. Chary, G. Govil, K. L. Liu, H. T. Miles, *Nucleic Acids Res.* **26**, 2981 (1998).
- [7] Y. G. Kim, J. Smith, M. Durgesha, S. Chandrasegaran, *Biol. Chem.* **379**, 489 (1998).
- [8] L. Kittler, A. Bell, B. C. Baguley, G. Lober, *Biol. Chem.* **379**, 519 (1998).
- [9] A. Sander, M. Ruess, S. Bereswill, M. Schuppler, B. Steinbrueckner, *J. Clin. Microbiol.* **36**, 2973 (1998).
- [10] E. Urbach, C. Schindler, S. J. Giovannoni, *FEMS Microbiol. Letts.* **162**, 111 (1998).
- [11] S. N. Ponnampalam, S. Elsen, C. E. Bauer, *J. Biol. Chem.* **273**, 30757 (1998).
- [12] R. Duan, W. Porter, S. Safe, *Endocrinology* **139**, 1981 (1998).
- [13] V. Martin, S. Ribieras, M. C. Rio, R. Dante, *Biol. Chem.* **379**, 409 (1998).
- [14] K. Mori, N. Ogawa, T. Kawahara, H. Yanagi, T. Yura, *J. Biol. Chem.* **273**, 9912 (1998).
- [15] C. S. Song, M. H. Jung, S. C. Kim, T. Hassan, A. K. Roy, B. Chatterjee, *J. Biol. Chem.* **273**, 21856 (1998).
- [16] M. Detapia, V. Metzler, M. Mougel, B. Ehresmann, C. Ehresmann, *Biochem.* **37**, 6077 (1998).
- [17] D. Pearce, W. Matsui, J. N. Miner, K. R. Yamamoto, *J. Biol. Chem.* **273**, 30081 (1998).
- [18] J. Lechner, T. Welte, W. Doppler, *Immunobiol.* **198**, 112 (1998).
- [19] K. Willwand, E. Mumtsidu, G. Kuntzsimon, J. Rommelaere, *J. Biol. Chem.* **273**, 1165 (1998).
- [20] C. H. Huang, Y. S. Lin, Y. L. Yang, S. W. Huang, C. W. Chen, *Molec. Microbiol.* **28**, 905 (1998).
- [21] M. Royo, M. A. Contreras, E. Giralt, F. Albericio, M. Pons, *J. Am. Chem. Soc.* **120**, 6639 (1998).
- [22] E. Ardini, G. Pesole, E. Tagliabue, A. Magnifico, V. Castronovo, M. E. Sobel, M. I. Colnaghi, S. Menard, *Molec. Biol. Evolut.* **15**, 1017 (1998).