Physica A 469 (2017) 244–249

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Autoantibody recognition mechanisms of MUC1

J.C. Phillips

Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, 08854, United States

HIGHLIGHTS

- Cost-effective early cancer biomarkers are based on p53 and mucin epitopes.
- Mucin molecular structure is dominated by partially glycosylated repeats.
- Analysis of hydrogen-bonded β strand propensities explains observed sensitivities.

ARTICLE INFO

Article history:
Received 13 July 2016
Available online 17 November 2016

Keywords:
Autoantibody
Cancer
Early detection
Mucin
Glycosylation

ABSTRACT

The most cost-effective blood-based, noninvasive molecular early cancer biomarkers are based on p53 epitopes and MUC1 tandem repeats. Here we use dimensionally compressed bioinformatic fractal scaling analysis to compare the two distinct and comparable probes, which examine different sections of the autoantibody population, achieving combined sensitivities of order 50%. We explain the experimental observation that glycosylation does not enhance, and can depress, the sensitivity of MUC1 tandem repeat biomarkers. We propose a possible supplementary MUC1 epitope in the SEA region outside the tandem repeats.

1. Introduction

The value of p53 printed 15-mer epitopes as biomarkers for seromic autoantibody paratopes, much more sensitive than complete 393 amino acid (aa) p53, was established in Ref. [1], using the ∼50,000 London patient database. In subsequent papers similar methods identified central MUC1 tandem repeats as also effective, and complementary [2,3]. We have developed thermodynamic scaling methods that utilize only parameter-free bioinformatic scales derived from the geometry of thousands of protein structures recently added to the Protein Data Base [4]. In Ref. [5] these scaling methods were tested against the p53 epitopes discovered in Ref. [1], with strongly positive results. Here we extend the analysis to MUC1 tandem repeats.

Because p53 is the most studied protein, it is an attractive subject for testing scaling methods, which appear to have universal features, including quantitative aspects of evolution not accessible to phylogenetic trees [4]. Between human and mouse, p53 has evolved with ∼80% BLAST similarity. Mucin forms multiple families, the ones most relevant to humans being MUC1 and MUC4. Much less is known about mucin than about p53, so we begin by reviewing its salient feature, the 20 aa tandem strict and peripheral variable repeats near the N terminal of MUC1 and the more variable 16 aa tandem repeats near the N terminal of MUC4. The number of these repeats varies usually from 30 to 90 in human individuals, but is only 16 in mouse [6]. Apparently evolution has expanded these repeats in humans, and [2,3] contain high-resolution data sensitive to the repeat structure.

Our scaling methods are focused on quantifying sculpted protein surface topology, making maximal use of modern bioinformatically derived amino acid scales. Modeling with earlier and less accurate scales achieved only limited successes,

---

E-mail address: jcphillips8@comcast.net.

http://dx.doi.org/10.1016/j.physa.2016.11.075

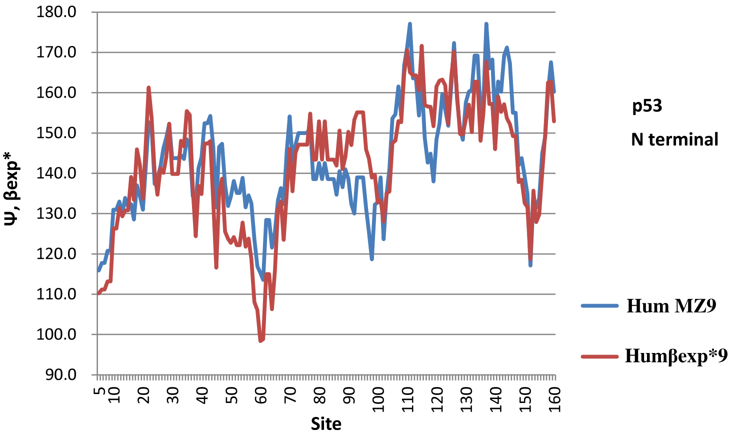

Fig. 1. The strongest signal for early cancer detection in Ref. [1] came from the 15 mers corresponding to 40–55 and 45–60. It was suggested in Ref. [5] that both signals could have been caused by a common 9 or 11 mer near 45–55 = LSPDDIEQWFT, with $\langle \Psi \mathrm{MZ} - \beta \exp^* \rangle \sim 10$. This region is strongly hydrophilic ($\langle \Psi \mathrm{MZ} \rangle \sim 137$). Bearing in mind the difference of 10 between $\langle \Psi \mathrm{MZ} \rangle$ and $\langle \beta \exp^* \rangle$, it is equally strong in its preference for exposed beta strands. It is possible that this epitope occurs as an early cancer signal, but does not occur with tumors, because it lies in the N terminal region, away from the DNA binding region.

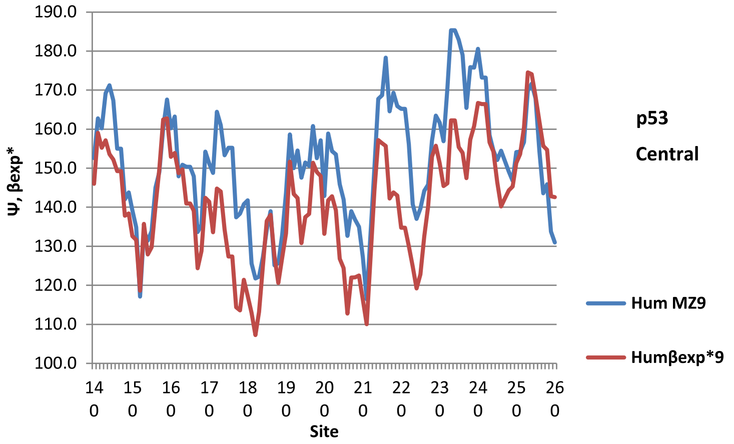

Fig. 2. Here is shown the central part of the DNA binding region 102–292. The strong 15-mer epitope for colon tumors is 165–180. This is the region that is most similar to 40–60 in Fig. 1, because it is both strongly hydrophilic (Ψ MZ), equally strong in its preference for exposed beta strands ($\Psi \mathrm{MZ} - \beta \exp^* \sim 10$). It is possible that this epitope occurs in tumors but not as an early cancer signal because it lies in the DNA binding region.

which have dampened interest in scaling approaches. Our work has encouraged us to suppose that an entire range of new and more accurate results can be achieved quite easily, using computationally trivial EXCEL macros and modern scales. The two scales used here are the hydropathic solvent-accessible surface area scale Ψ based on fractals [4,7], and the beta strand exposed scale $\beta \exp$ [8,9]. These two scales are strongly anticorrelated ($r \sim -0.9$), and their unweighted averages are $\langle \Psi \rangle = 155$ and $\langle \beta \exp \rangle = 145$. We plot $\beta \exp^* = 300 - \beta \exp$ to make the intrinsic surface similarities more obvious.

We are concerned here primarily with detecting interactions of p53 epitopes or mucin fragments with autoantibody paratopes. We previously found good results for p53 epitopes [1,5] using the beta strand exposed scale $\beta \exp$ averaged over a sliding window wave length $W = 9$, which is also the lower limit of the fractal Ψ range [7]. By using this value of W, we achieve maximum resolution. In addition, we have examined the correlation r between Ψ and $\beta \exp^*$ for the canonical 1VTSAPDTRPAPGSTAPPAHG20 tandem repeat unit of MUC1, and found a remarkable result. For $W = 1$ (no averaging), $r = 0.72$, but this value increases to 0.90 (0.01) for $W = 9$ and 11, where it plateaus at the same value as the correlation r of unweighted scales. The effect of using $W = 9$ for mucin tandem repeats is to recover the strong interscale correlation contained in the original bioinformatic scales. Given the remarkable physiology and repeat structure of mucin, this conclusion is less surprising than it might have been for any other protein. It also strengthens the universality of the bioinformatic results [7,8]. Much lower r values are found for other parts of MUC1, such as the conformationally stressed cleavage domain SEA [9].

We review the p53 results in Figs. 1 and 2, which separate p53 into two parts for clarity. The most effective epitopes are associated with segments that have $\beta \exp^*$ below Ψ by about 10 scale units. Specifically the 40–60 epitope (which could contain a 49–57 9-mer nucleus [5]) also is near the deepest hydrophilic minimum (most exposed, and most flexible) sites of p53.