Evolution of linear motifs within the intrinsically disordered and globular domains of the papillomavirus E7 oncoprotein

Lucía B. Chemes¹, Juliana Glavina², Julián Faivovich³, Leonardo G. Alonso¹, Cristina Marino-Buslje⁴, Gonzalo de Prat Gay¹, Ignacio E. Sánchez²

1Protein Structure-Function and Engineering Laboratory and 2Structural Bioinformatics Laboratory. Fundación Instituto Leloir and IIBBA-CONICET, Buenos Aires, Argentina. 3Protein Physiology Laboratory, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina. 4División Herpetología, Museo Argentino de Ciencias Naturales-CONICET, Buenos Aires, Argentina.

INTRODUCTION

Many protein functions can be described in terms of linear sequence motifs of less than five function-determining residues, which are often found within intrinsically disordered domains¹¹. Linear motifs pose new challenges to the evolutionary biologist and are only beginning to be understood. These motifs appear or disappear with only a handful of point mutations and are thought to evolve rapidly, presumably because a flexible structural context facilitates linear motif evolution and function¹²,². On the other hand, disordered regions show amino acid substitution patterns and structural conservation that are clearly different from those of globular domains³⁴. The relationship between disorder and linear motif evolution is only beginning to be explored. Evolutionary studies of linear motifs only recently started to consider the sequence and structural context⁵⁶, while studies focusing on disordered regions as a whole usually neglect the linear motifs they contain³⁷. Moreover, linear motifs of a protein are considered individually, disregarding potential functional and evolutionary coupling. This state of the art calls for the development of new paradigms for studying the evolution of linear motifs within disordered and ordered regions.

We have chosen the papillomavirus E7 oncoprotein as a model to study sequence conservation and linear motif evolution. The wealth of clinical, biochemical and sequence information makes this protein a useful system for molecular evolution studies⁶⁷. E7 presents two structural domains, a disordered N-terminal domain (E7N) and a globular C-terminal domain (E7C). The comparison between a disordered and a globular domain provides an excellent opportunity to study the interplay between order and disorder in sequence and linear motif evolution. Here, we focus on the functional linear motifs in the papillomavirus E7 oncoprotein and how their structural environment modulates their evolution. Functional linear motifs in viral proteins are likely candidates for convergent and adaptive evolution because they can appear or disappear in very short timescales with only a handful of point mutations²⁷, but this widely held hypothesis has not been tested extensively. The small number of genes in papillomaviruses should facilitate the search for relationships between genotype and phenotype at the level of linear motifs.

RESULTS

Sequence conservation and co-evolution in E7N and E7C

We used over 200 sequences corresponding to the E7 protein from known papillomavirus types to construct sequence alignments for the E7N and E7C domains. We used the information content of each position in the alignment as a measure of conservation, and an algorithm based on mutual information to identify pairs of coevolving residues. E7N and E7C showed similar degrees of conservation (E7N 2.4 ± 1.0 bits, E7C 2.1 ± 1.2 bits), indicating that a lack of globular structure does not necessarily lead to a lower degree of sequence conservation. E7C conservation and co-evolution patterns were best explained in terms of its globular and homodimeric structure, while conservation and co-evolution in E7N were best explained in terms of densely packed linear motifs separated by variable linkers. We found pairs of co-evolving residues within each domain as well as pairs whose residues belonged each to a different domain, indicating that the evolution of the E7N and E7C domains was not entirely independent.
Variability and association of E7 linear motifs

We studied the variability in the linear motif repertoire for different E7 proteins, including eight linear motifs located in the E7N and E7C domains. Each motif was given a definition based on its conserved sequence pattern, and bioinformatics tools were used to score the presence/absence of each motif in all E7 protein sequences. The motif repertoire was then represented superimposed on a phylogenetic tree of papillomaviruses. The different linear motifs showed different abundance and distribution patterns, and most motifs were not homogeneously distributed on the papillomavirus phylogenetic tree. Six of the eight motifs were almost fully present in the alpha genus, which contains all genital human papillomavirus types, and only partially present in most of the other supertaxa. This shows that the motifs defined from detailed studies on only a few types and the corresponding viral activities cannot be extrapolated to all papillomaviruses. Interestingly, statistical association tests revealed co-occurrence for the four most prevalent E7N motifs, suggesting that they form functional and evolutionary units. These results suggest E7 motifs do not evolve independently, perhaps due to coupling of their functions.

Convergent evolution of E7 linear motifs and association with phenotype

The uneven distribution of E7 linear motifs in the phylogeny suggested that they had appeared or disappeared more than once during papillomavirus evolution. We tested for this hypothesis by reconstructing the evolutionary history of the eight motifs using maximum parsimony. For most motifs, evolutionary changes occurred in several papillomavirus clades. Most motifs showed multiple independent disappearance events in different branches of the tree. Some motifs appeared only once during papillomavirus evolution, indicating common ancestry of all extant motifs. However, most motifs appeared independently several times during papillomavirus evolution, indicating that some instances of the motif did not have common ancestry. Multiple independent appearances of the motifs in E7 indicated that they were prone to convergent evolution. The evolutionary pattern of the different motifs did not depend on whether they were located in a globular or disordered domain, indicating that the evolutionary plasticity of linear motifs in E7 is not strongly determined by the presence of local disorder.

Different papillomavirus species are associated to phenotypes such as tissue tropism or type of lesion, which are well grouped in the papillomavirus phylogenetic tree. The appearance or disappearance of linear motifs in specific branches of the phylogenetic tree coincided with changes in phenotype within that branch. Statistical tests confirmed the association between several linear motifs and specific phenotypes, providing strong evidence that the appearance/disappearance of the motifs contributed to adaptive evolutionary events leading to changes in phenotype.

CONCLUSIONS

The present study confirms the widely held hypothesis that linear motifs are able to evolve rapidly, showing that the eight motifs in E7 have changed multiple times during papillomavirus evolution. However, the evolutionary plasticity of the different motifs did not depend on whether they were located in a globular or disordered region. This result indicates that motif evolution in E7 is not strongly determined by the presence of local disorder, contrasting with the current view that a flexible structural context should facilitate linear motif evolution. Sequence conservation, co-evolution, and association and evolutionary analyses show that the linear motifs in E7 do not evolve independently but instead show strong functional and evolutionary coupling. Finally, multiple independent appearances of several motifs during papillomavirus evolution provide direct evidence for convergent evolution, which may play an adaptive role as shown by correlation with phenotype. This case study of motifs within the E7 viral protein confirms previous hypotheses that linear motifs in viral proteins are likely candidates for convergent and adaptive evolution.