3D-Structure Prediction of the Modular Protein Sialoadhesin Using a Multi-step Modelling Strategy

Corrada D[1,2], D'Ursi P[1], Botti S[3], Luperini A[4], Milanesi L[1], Rovida E[1]

[1]Institute for Biomedical Technologies – National Research Council (ITB-CNR), Segrate (MI)
[2]Department of Biomedical Science and Technologies, University of Milan, Segrate (MI)
[3]Parco Tecnologico Padano – Centro Ricerche e Studi Agroalimentari (PTP-CERSA), Lodi
[4]ENEA TER ENE-SIST, Ispra (VA)

Motivation

Many complex pathways involved in cell-cell recognition and in signal trasduction of cellular trafficking primarily rely on glycan-receptor interaction. Most of these receptors belong to the superfamily of immunoglobulins (IgSF) which includes the Sialic acid immunoglobulin lectines (Siglec). Siglec1 (or Sialoadhesin, Sn) is a type I membrane glycoprotein of about 1600 residues length, consisting of 17 Ig-like domains whose reciprocal fold has not been clarified. The protein is present in different species and is implicated in infection by viruses such as rhinovirus, PRRS virus, HIV virus. Sequence is highly conserved among mammalian suggesting also a high structural conservation. The knowledge about catalytic domain (SnD01) is well documented, also in terms of crystallographic structures. In a project aimed to investigate the genetic components of pig natural immunity segregating in commercial herd, we studied the genetic variability of Siglec1. We have identified 35 non-silent SNPs in Siglec1 coding regions. In this work we propose a structural model for the first eight domains (SnD01 to Snd08) of sialoadhesin with the purpose to locate the spatial position of the mutated residues and to gain new insight into functional aspects of the protein.

Methods

Strand positions along the amino acid sequence (UniProtKB, Q7YRQ7) were defined by construction of a consensus secondary structure profile. The threading methodology implemented by BioInfoBank Meta Server was applied to collect the templates showing the best ranked alignments by the Sn sequence. 3D structures of Siglec1 domains were generated by the following protocol. In the first step, we have produced 250 alternative structural conformations from each alignment: each structure were analyzed and ranked on the basis of energy scoring functions. The top ranked structure for each template was investigated in terms of energy profile to find the most disordered regions.. In the second step, we have performed a loop refinement in order to re-model these regions: for each raw model, we have produced 50 further fine conformations. Top ranked models of Sn domains were validated by visual analysis and ProCheck submission.

Results

Due to the complexity of domains clustering in the protein, the modelling procedure required the use of multiple templates. Homology modelling technique was applied to realize structural alignments for domains SnD01 and SnD02. For domains from SnD03 to SnD17 we could not find any entry that matched our sequence with an identity higher than 15%: for this portion of the protein, we applied a threading approach. The whole sequence as well as subsets of variable length (2,3, and 4 domains) were submitted to BioInfoBank Meta Server: the best ranked alignments were found for the input sequences corresponding to couple of domains (SnD03/SnD04; SnD05/SnD06; etc.). We therefore obtained a set of models, each corresponding to pairs of SnD domains from SnD01 to Snd08. The modelled portions were joined by overlapping their terminal extensions (see figure attached). We have also performed surface analyses in order to identify the presence of cavities and predict putative interaction sites. The overall model allowed us to localize 15 of the 35 SNPs identified in pig Siglec1. Some of them were found on interdomain loops, and they could fill a role in conformational changes of the global structure; others SNPs were positioned around some of the predicted cavities, and may be involved in defining the functional shape of putative binding sites. The functional meaning of the substitutions were evaluated considering the structural environment as well as the evolutionary conservation of the position. In vivo or in vitro data will be required to correlate SNPs induced alterations of biological activity with the structural consequences of the variation.

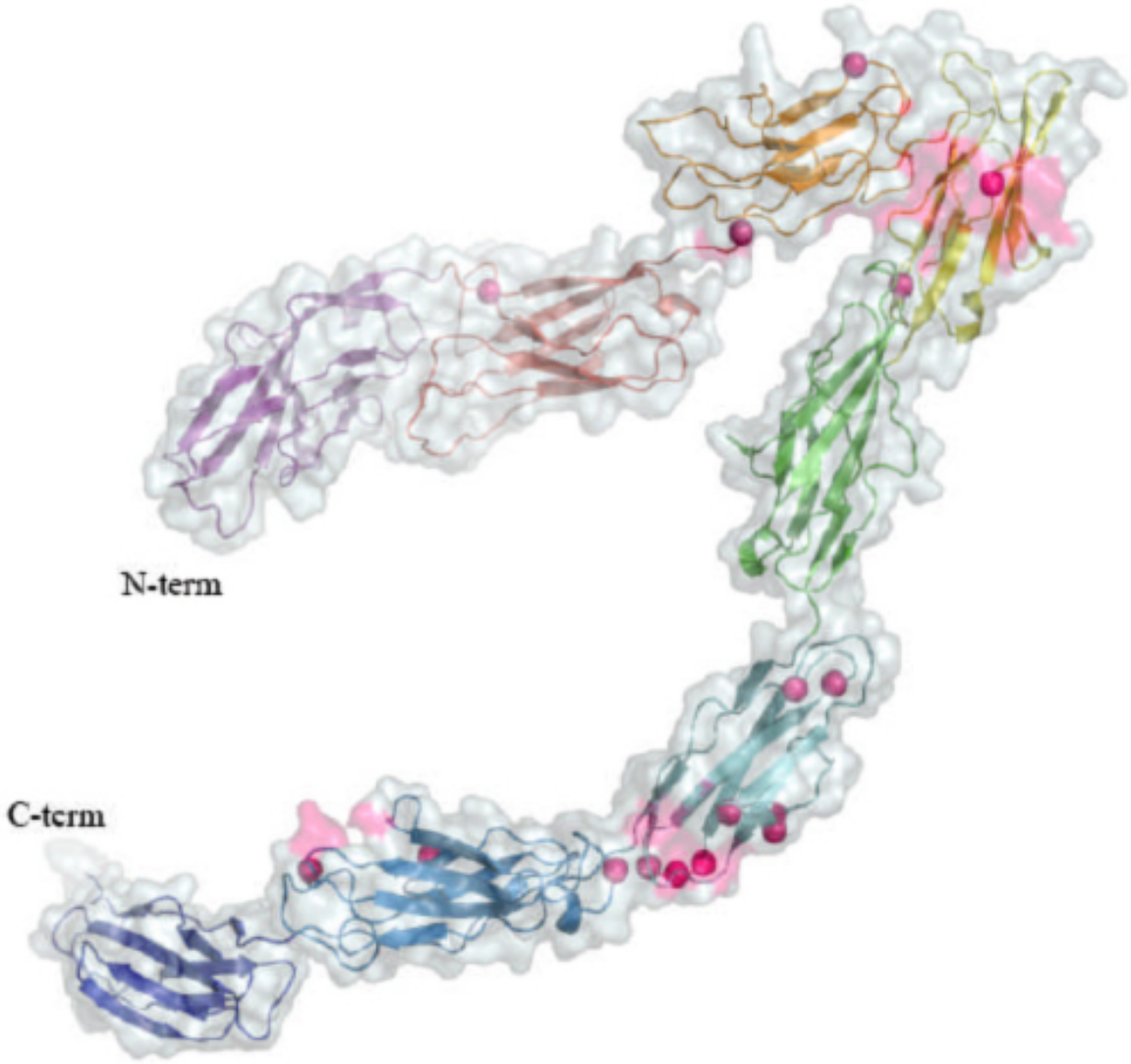

Figure 1: an overview of the first 8 N-terminal Sn domains joined together through structural overlaps of terminal extension. Positions of SNPs are showed as spheres; predicted cavities nearby SNPs are highlighted on molecular surface.

Contact : dario.corrada@itb.cnr.it