Identification and in silico characterization of double histone fold domains in Cca3 and "Similar to Cca3" proteins

Greco C. , Sacco E. , Vanoni M. , De Gioia L.

Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan (Italy).

Motivation

Histone folds are structural elements that are able to form dimers by means of tight interactions between hydrophobic surfaces. Normally, a histone fold is composed by a long alpha-helix flanked by two or three shorter helices. In the nucleosome core particle, two pairs of H2a-H2b and H3-H4 histone heterodimers assemble together, giving rise to a disk-like octamer upon which DNA rolls up. The publication of the X-ray structure of the prokaryotic histone from Methanopyrus kandleri highlighted a novel protein fold, which is originated by the assembly of two consecutive histone folds included in the same peptide chain. More recently, the publication of the X-ray structure of the amino-terminal domain of hSos1 showed that also this protein module assumes a similar fold, dubbed the histone pseudodimer and here also referred to as "double histone fold". In fact, the evolutionary relationship between the H2a histone and the domain spanning the protein sequence 96-190 of hSos1 had been already disclosed, due to high sequence similarity between the two domains. However, the first histone-like domain of hSos1, spanning the protein portion 6-95, does not show evident sequence similarity with histones (Sondermann et al, 2003). Moreover, it is unknown whether the double histone fold can be found in other protein families. In view of this, we initiated an in silico study aiming at the identification of other proteins characterized by a sequence that is compatible with the double histone fold.

Results

The alignment between the domain spanning the hSos1 protein portion 22-95 and the histone H2b was analysed to evaluate possible remote homology. Remarkably, it turned out that the H2b domain and the corresponding hSos1 domain show a significant similarity (22.1% identity). The analysis of the alignment highlights a strict conservation of a group of hydrophobic residues; the conservation of this feature is crucial for the formation of tight interactions between the two histone folds and for the correct folding of this domain. Prompted by the above observations we searched for other proteins that could be characterized by the peculiar double histone fold observed in hSos1. Submission of the N-terminal protein sequence of hSos1 (residues 1-190) to Psi-Blast did not highlight proteins sharing a significant sequence similarity with the probe sequence, in agreement with previous observations (Sondermann et al, 2003). However, the existence of homology between Sos proteins and histone H2b (see above) led us to the application of a different strategy for the search of remote homologues, based on the submission to Psi-Blast of a "chimeric" sequence obtained linking the protein regions corresponding to the H2b (residues 38-114) and H2a (residues 1-119) monomers of the human histone dimer. This search gave statistically significant hits corresponding to the rat Cca3 protein and its homologues from many other organisms; we also identified two other proteins from rat and chicken tagged as "similar to Cca3" (Cca3S), and their homologues from Homo sapiens, Xenopus tropicalis and Tetraodon nigroviridis. Further analyses based on three different secondary structure prediction servers (PHD, J-pred and Psi-pred) and on the fold recognition

metaserver 3D-jury confirmed that the aminoacidic sequences of all these proteins are compatible with the double histone fold. In Cca3 proteins, an amino terminal module of variable length precedes the region of homology to histones, which is followed by ankyrin repeats and a POZ domain; as for Cca3S proteins, they are significantly shorter than Cca3, and do not include ankyrin repeats or POZ domains. However, Cca3S proteins show a significant sequence similarity in the region corresponding to the putative double histone fold, and in the upstream amino-terminal module. A homolgy model of the histone pseudodimer included in rat Cca3 was constructed; a subsequent study based on sequence alignments and on the calculation and analysis of electrostatic potential maps revealed that the histone pseudodimers included in Cca3 and Cca3S proteins, but not those present in Sos proteins, could retain the ability of mediating protein-DNA interactions. Previous literature data indicated that the prokaryotic histone pseudodimer could have a role in chromatin packaging, while the double histone fold domain in Sos proteins is known to be implicated in the mediation of intramolecoular interactions. The results of our in silico characterization of Cca3 and Cca3S proteins suggest that the double histone fold could also be involved in mediating the interactions between DNA and multidomain proteins. The analysis of new genomic sequences and the availability of more sensitive computational tools will allow to verify if other protein families can be included in the group of histone pseudodimer containing proteins.

Contact email: claudio.greco@unimib.it