The role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins

Piero Fariselli e Rita Casadio

Universita' di Bologna - via Irnerio 42, 40126 Bologna

A neural network-based predictor is trained to distinguish the bonding states of cysteine in proteins starting from the residue chain. Training is performed using 2452 cysteine-containing segments extracted from 641 non homologous proteins of well resolved 3D structure. After a cross-validation procedure efficiency of the prediction scores as high as 72% when the predictor is trained using protein single sequences. The addition of evolutionary information in the form of multiple sequence alignment and a jury of neural networks increase the prediction efficiency up to 81%. Assessment of the goodness of the prediction with a reliability index indicates that more than 60% of the predictions have an accuracy level greater than 90%. A comparison with a statistical method previously described and tested on the same database shows that the neural network-based predictor is performing with the highest efficiency.