COMPUTATIONAL PROTEOMICS:

FUNCTIONAL SITES ANALYSIS AND RECOGNITION IN 3-D PROTEIN STRUCTURE

N. A. Kolchanov

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

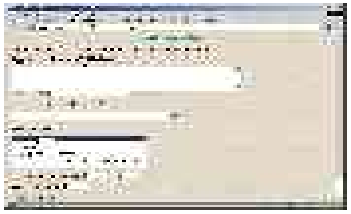
ИНСТИТУТ ЦИТОЛОГИИ
И ГЕНЕТИКИ

Tools for computer structural biology

PDBSite: a functional site database

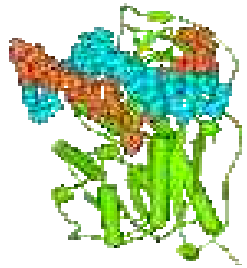
PDBSiteScan: a program for functional site recognition

PDBSiteComplex: a program for molecular complex reconstruction



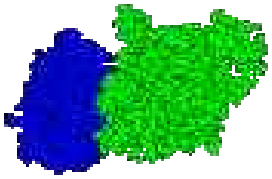
WebProAnalyst: a program for quantitative structure-activity relationships analysis in protein families

Protein function annotation



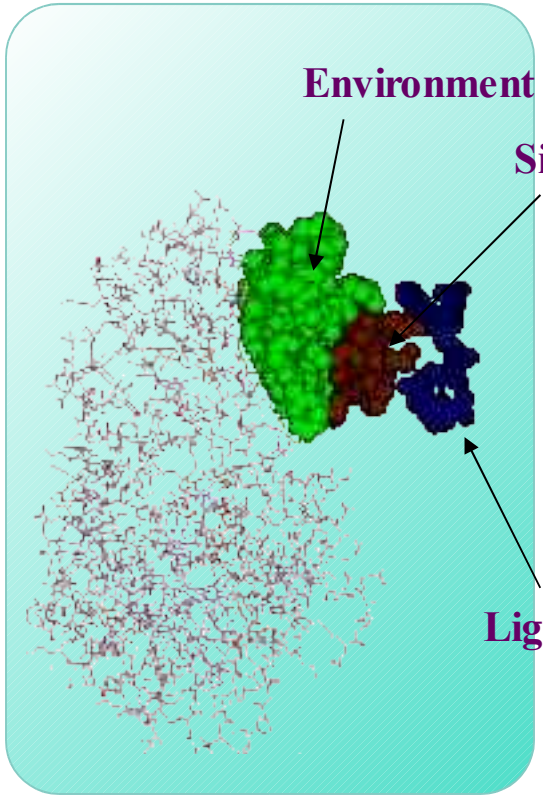
Molecular evolution

Search for drug targets



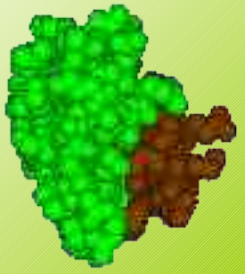
Data on spatial structure and main features of functional sites of proteins and their ligands are accumulating in the PDBSite and PDBLigand databases

http://wwwmgs.bionet.nsc.ru/mgs/gnw/pdbsite/

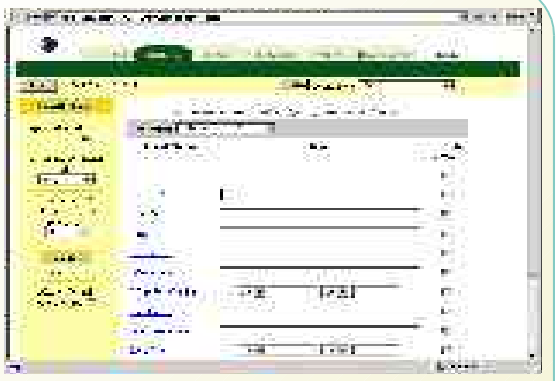


Annotated proteins from PDB are data sources

PDBSite database

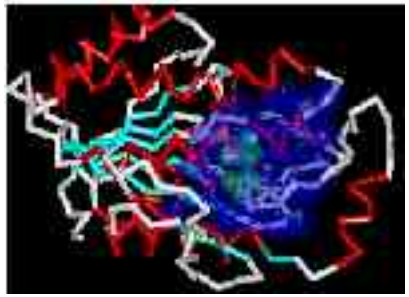


PDBLigand database



Ivanisenko et al, (2005) Nucleic Acids Res.

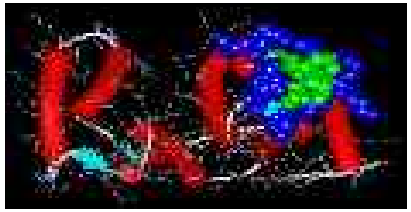
The PDBSite database contains



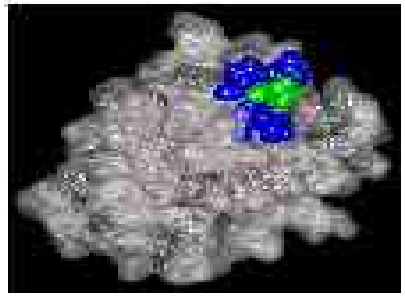
Active (1,300)

Drug binding (50)

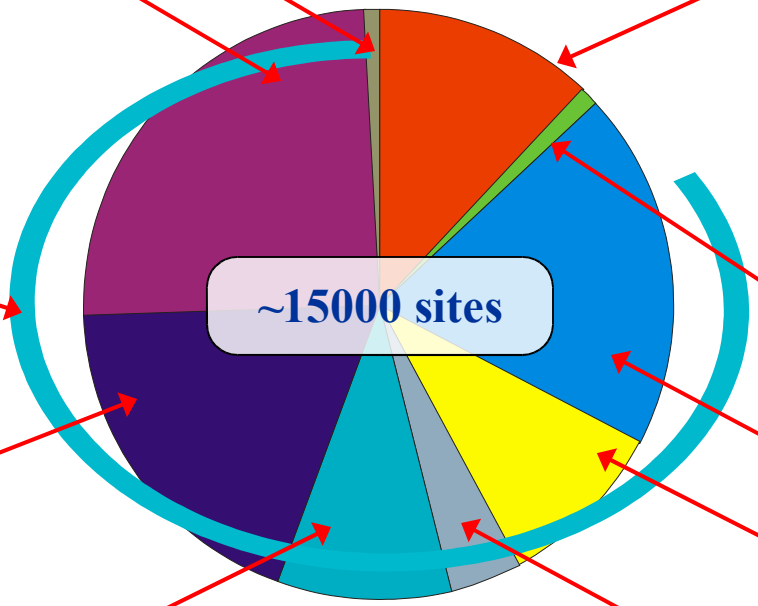
Protein-DNA (2,700)



Binding sites



Posttranslational modifications (100)



Organic ligands (2,100)

Protein-RNA (2,000)

Metal ions (1000)

Protein-protein (1,000)

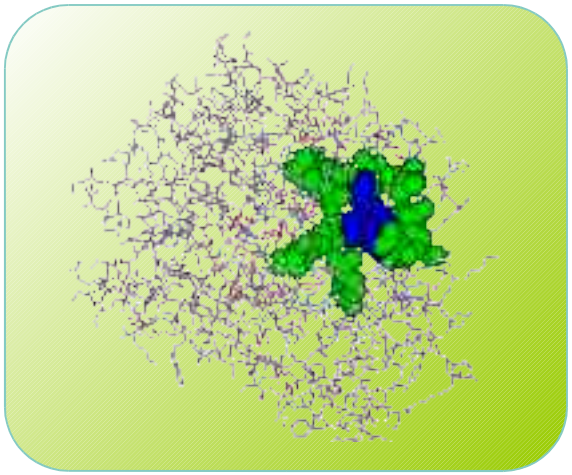
Inorganic ligands (400)

Ivanisenko et al, (2004)
Nucleic Acids Res.

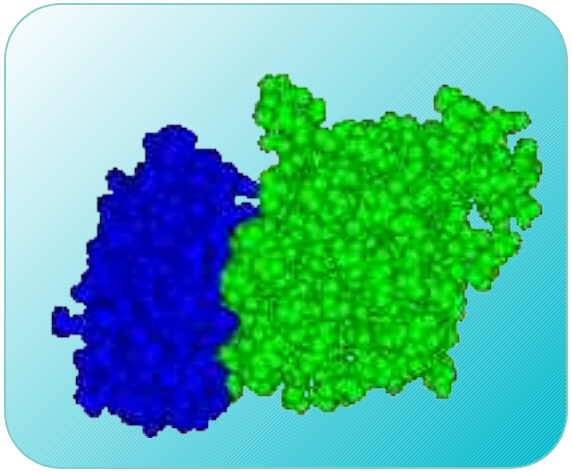
PDBSiteScan: a program for the recognition of functional site

http://wwwmgs.bionet.nsc.ru/mgs/systems/fastprot/pdbsitescan.html

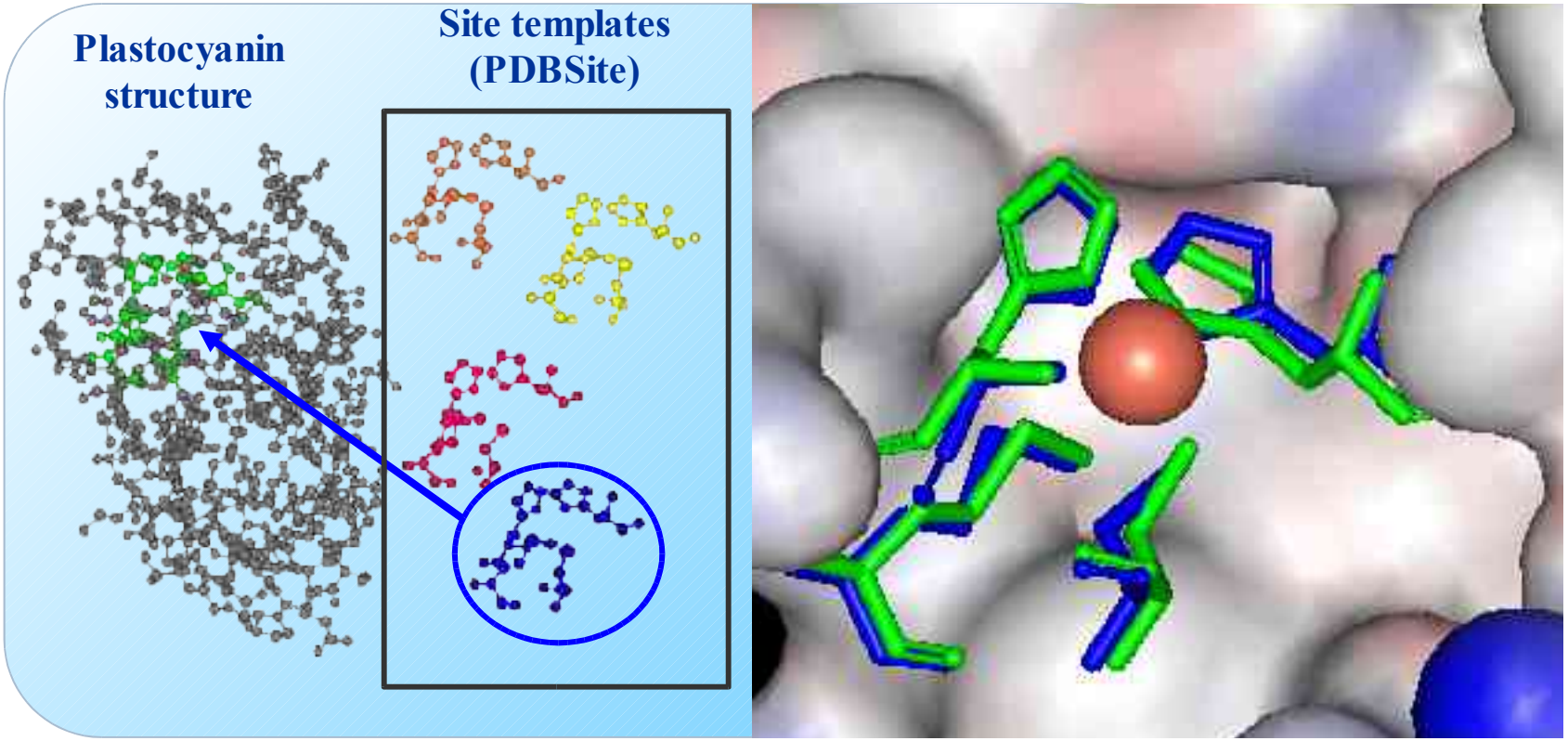
Functional site recognition



Reconstructio n of site-ligand complexes

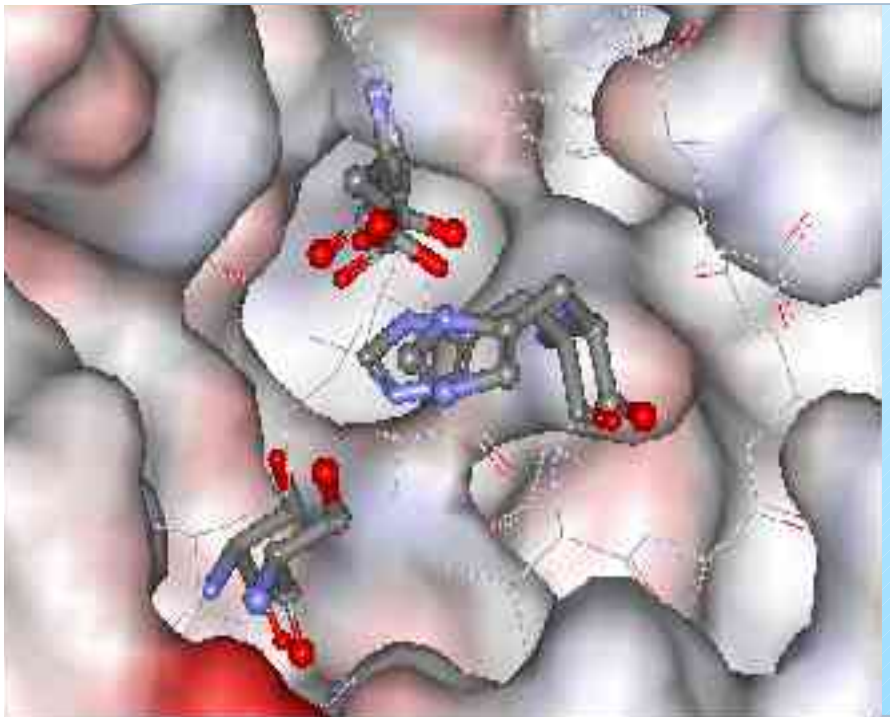


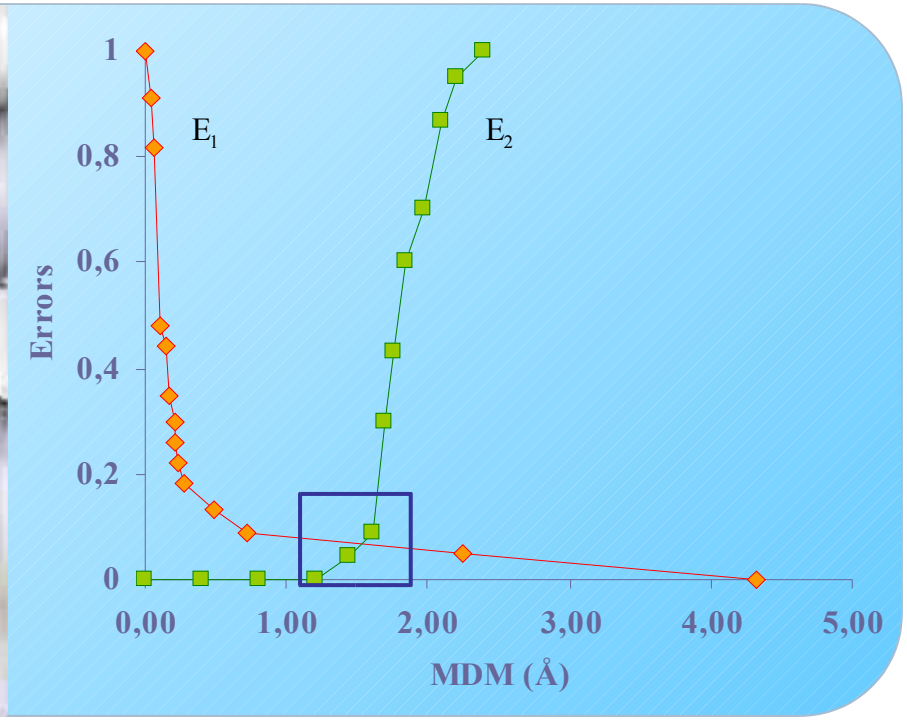
Example. Search for copper binding site in plastocyanin (PDB ID 1BXU)



The residues of the recognized site in plastocyanin are in green, those of template site from the PDBSite database (ID 1B3ICU) are in blue. Orange ball highlights copper ion.

Accuracy estimation for catalytic center recognition in hydrolase superfamily



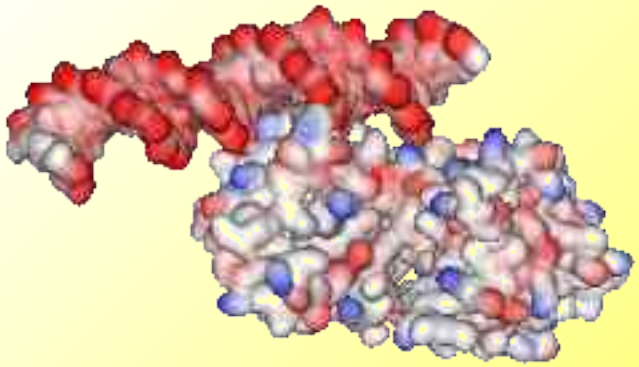


MDM – maximum distance mismatch between site template and protein fragment.

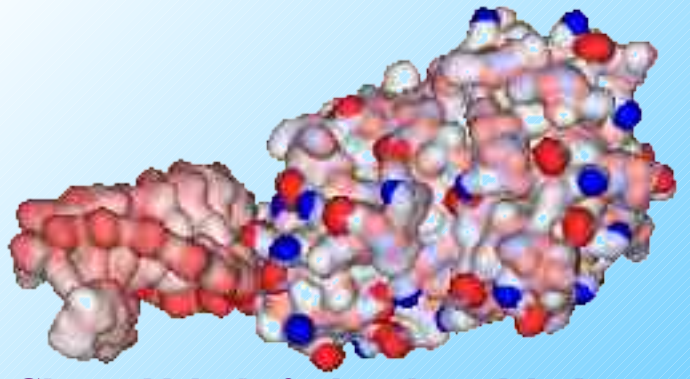
E_1 – type I error (underprediction)

E_2 – type II error (overprediction)

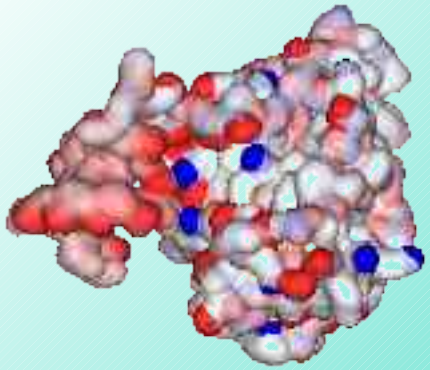
PDBSiteComplex: tools for reconstruction of interactions between proteins and DNAs/RNAs



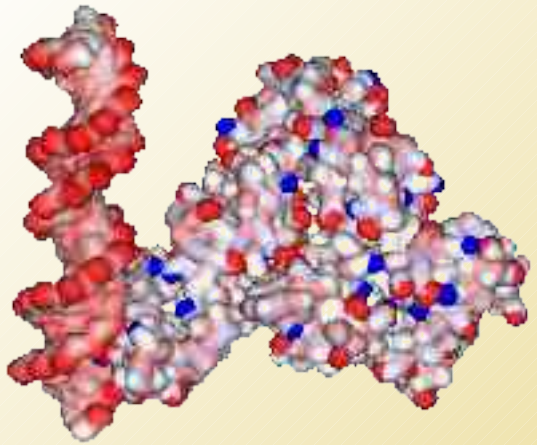
Phosphoglycerate kinase 1 - DNA



Glyceraldehyde-3-phosphate-dehydrogenase (GPD) - RNA



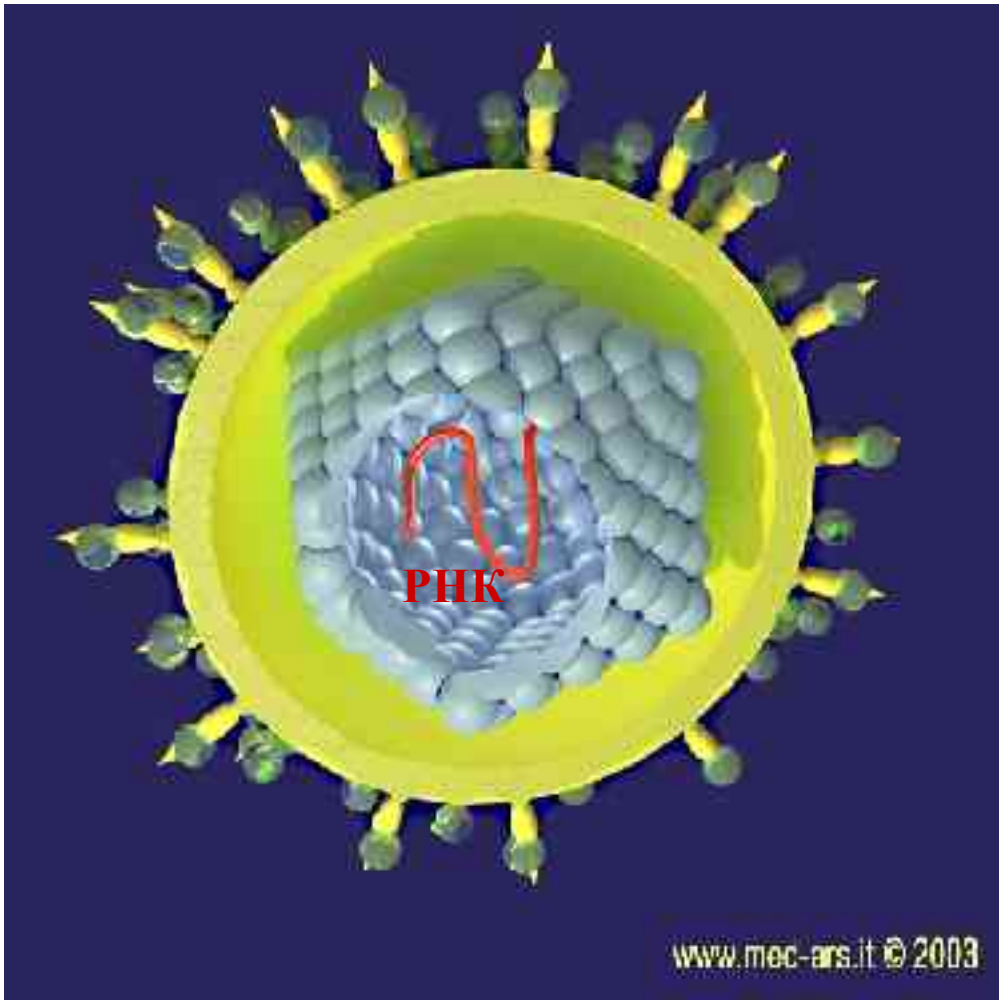
N-terminal domain of heat shock protein HSP90 - RNA



Serum albumin - DNA

Экспериментальные данные получены в ИХБиФМ СО РАН, Chelobanov et al, 2004

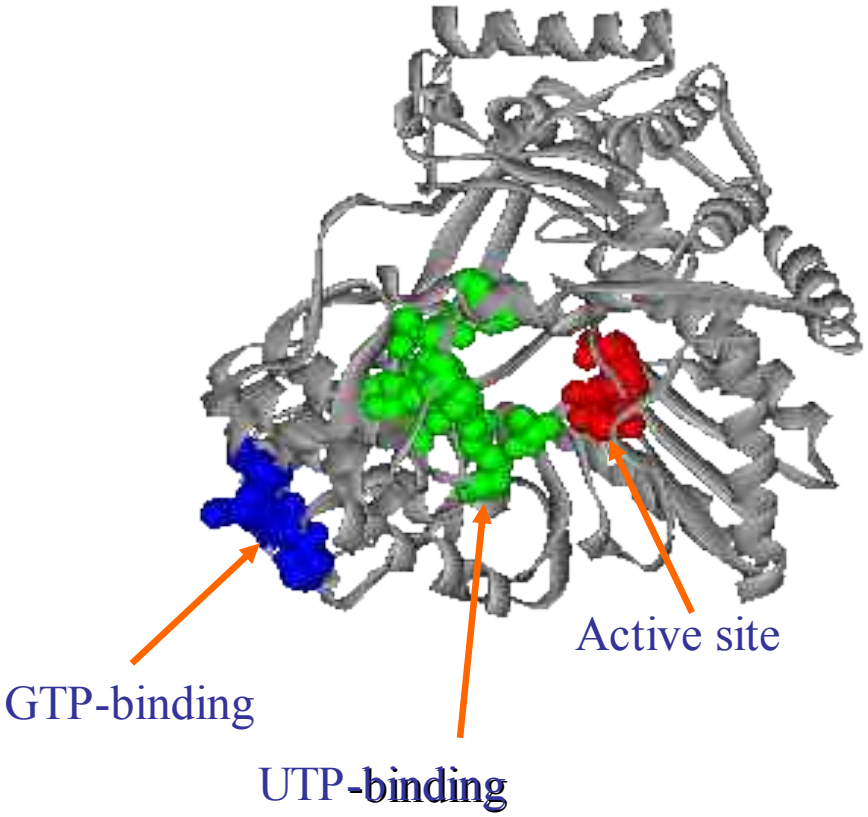
Hepatitis C virus: search for potential antiviral drug targets



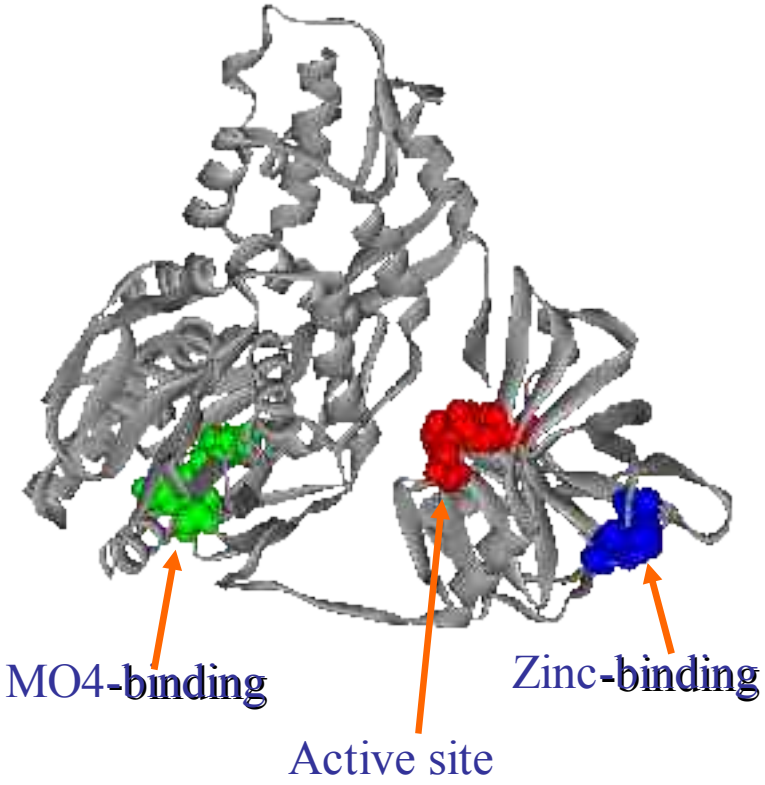
Hepatitis C virus is a member of flavivirus family. It enters the organism in a manner similar to HIV, affects liver. The chronic infection is protracted for 10-15 years, causes liver cirrhosis, provokes cancer, suppresses immune system. The number of virus C hepatitis infected humans in Russia is estimated as 2 millions.

Potential targets for anti-hepatitis C virus drugs

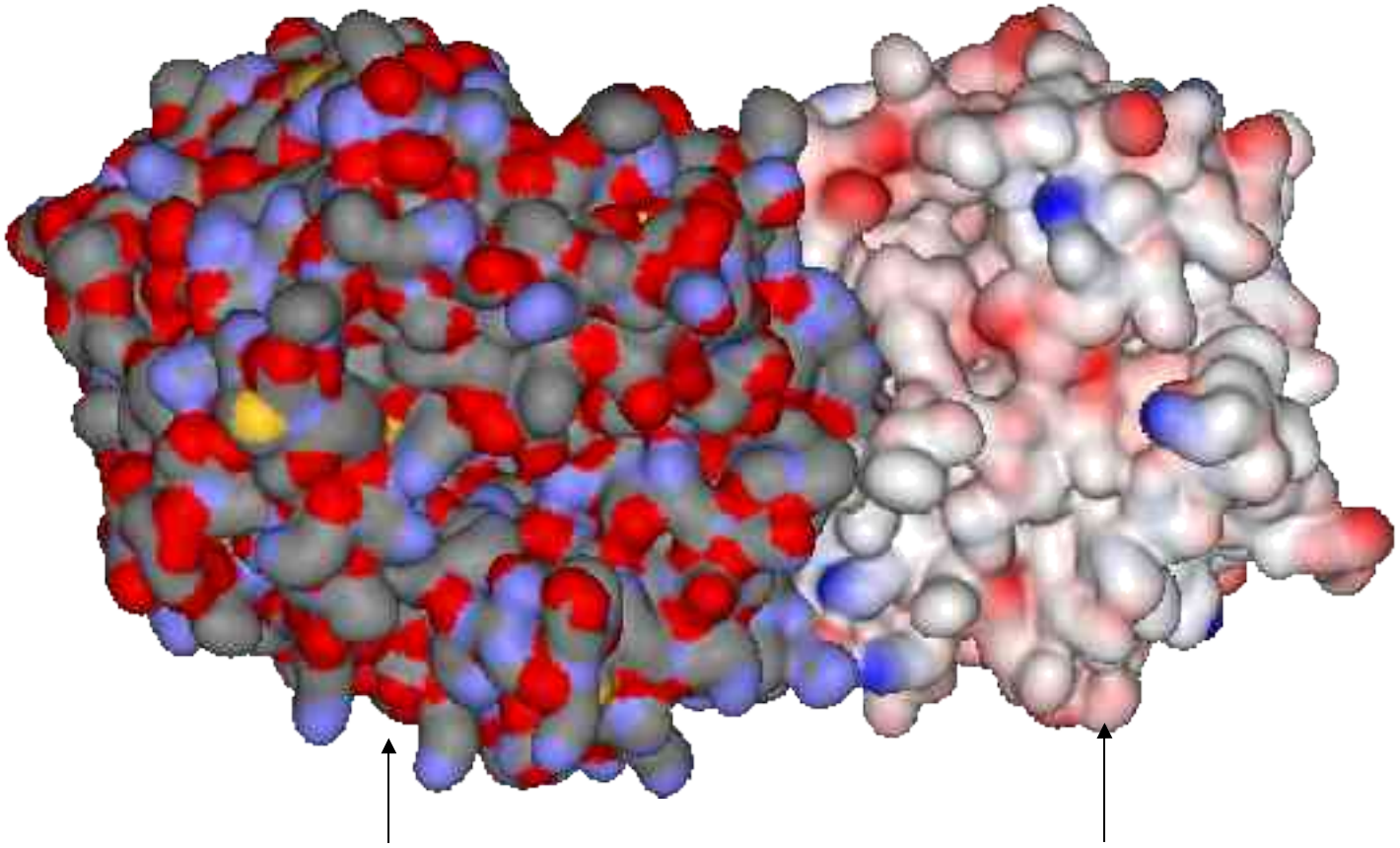
HCV NS5B transferase



HCV NS3 hydrolase

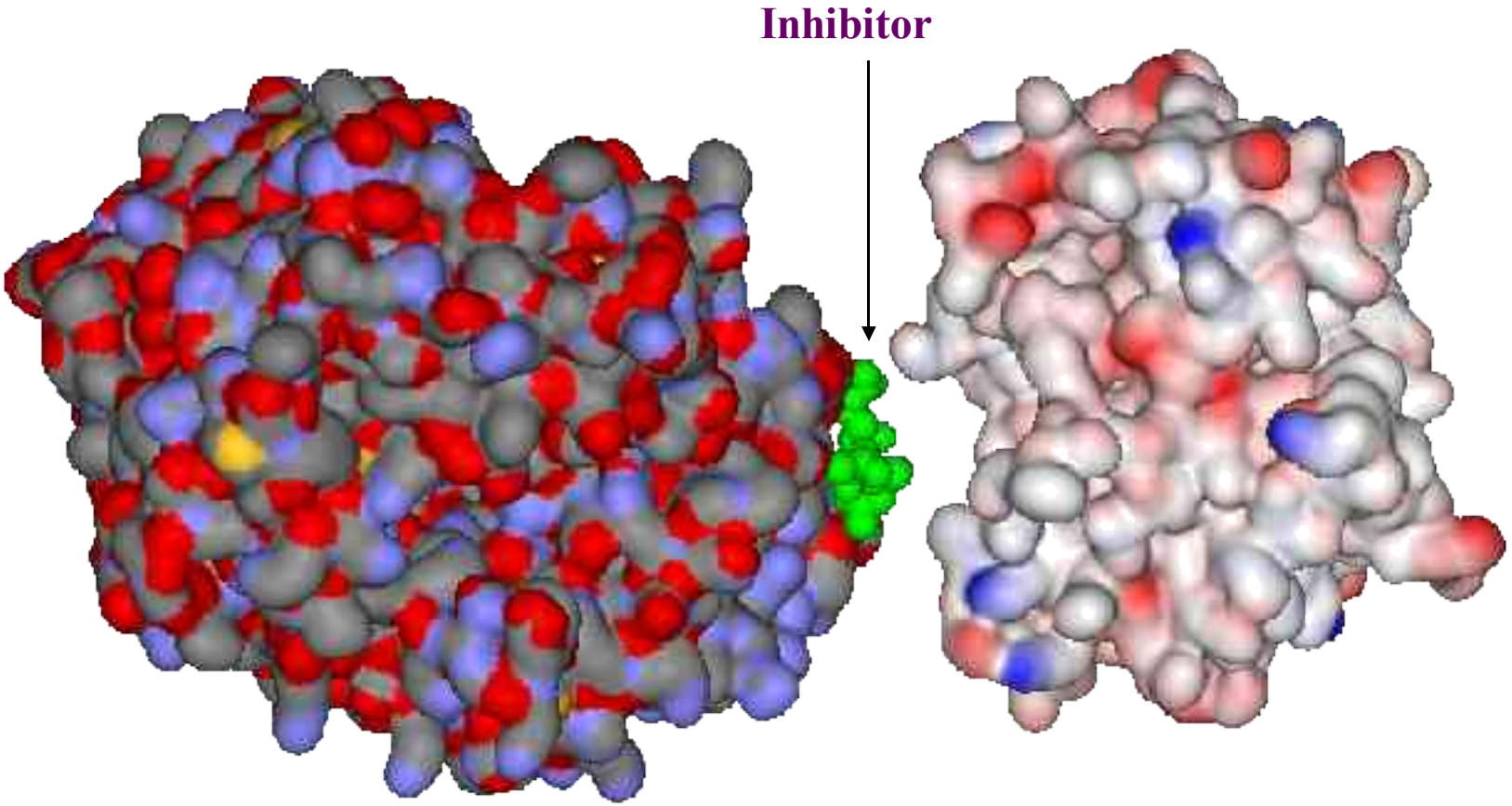


Potential targets for antiHCV drugs



Predicted complex formed by HCV NS5B protein with human NTF2 protein providing transport of proteins to cell nucleus

Potential targets for antiHCV drugs



Inhibitor of the formation of NS5B-NTF2 complex prevents NS5B transport into cell nucleus.

WebProAnalyst: a program for analysis of quantitative structure-activity relationships in homologous protein families

http://wwwmgs.bionet.nsc.ru/mgs/programs/panalyst/

Multiple alignment of protein sequences

Activities

Sliding window

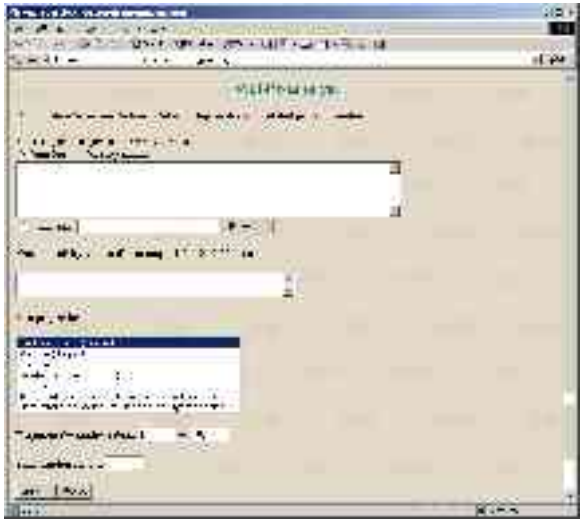
Protein 1
Protein 2
Protein 3
Protein 4
Protein 5

y1
y2
y3
y4
y5

Site

Calculation of amino acid physicochemical properties

x1
x2
x3
x4
x5



Search for significant relation
Y = F(X)

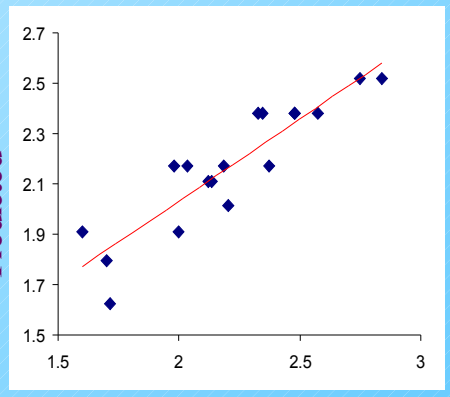
Protein activity

Site physicochemical properties

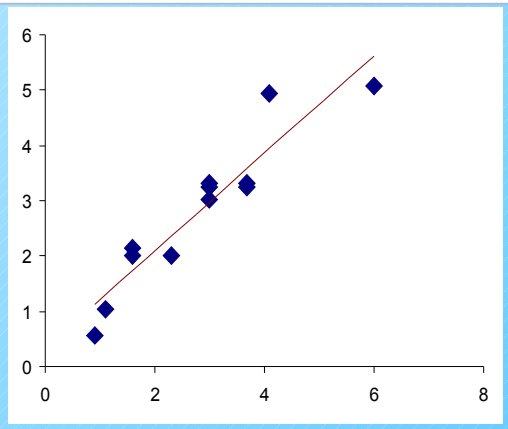
ИНСТИТУТ ЦИТОЛОГИИ
И ГЕНЕТИКИ

Examples. Quantitative structure-activity relationships in protein families

Relation between disintegrin capacity to inhibit platelet aggregation and site properties (charge and hydrophobic moment).



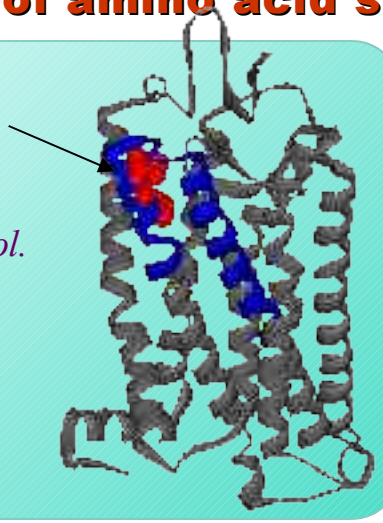
Relation between peptide antimicrobial activity and site hydrophobic moment.

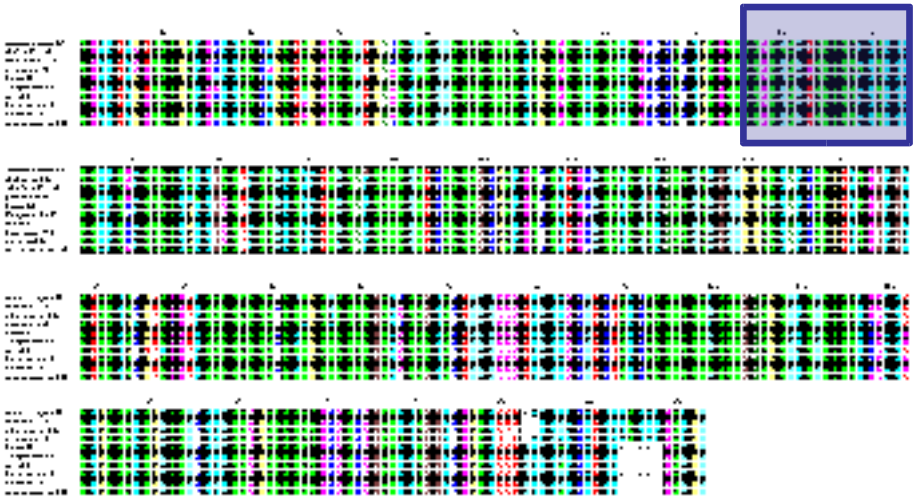


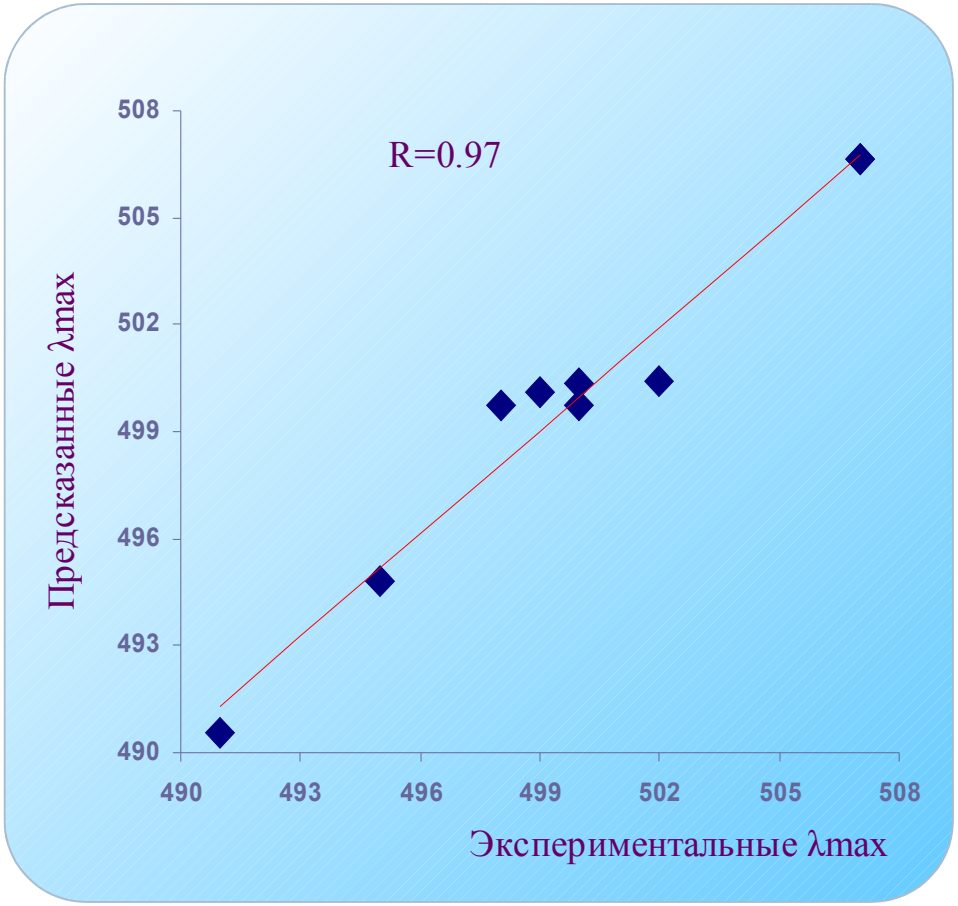
RODOPSIN PROTEIN IS RESPONSIBLE FOR LIGHT RECEPTION IN THE VISIBLE SPECTRUM

We have developed the program for RODOPSIN λmax prediction on the basis of amino acid sequence of this protein

Residues affecting shift of wavelength. The results agree with those of *Briscoe et al, Mol. Biol. Evol. (2001).*







Predicted λmax: **Y= 506.7**

Measured λmax: **Y= 508**

Regression equation:

$Y=15.784*X_1-467.266*X_2-37.661$

X_1 – mean for site isoelectric point (Bogard)

X_2 – hydrophobic moment (Eisenberg)

ИНСТИТУТ ЦИТОЛОГИИ
И ГЕНЕТИКИ

Prediction of the visible spectrum of archosaur sight (λmax for ancestral archosaur rodopsin)

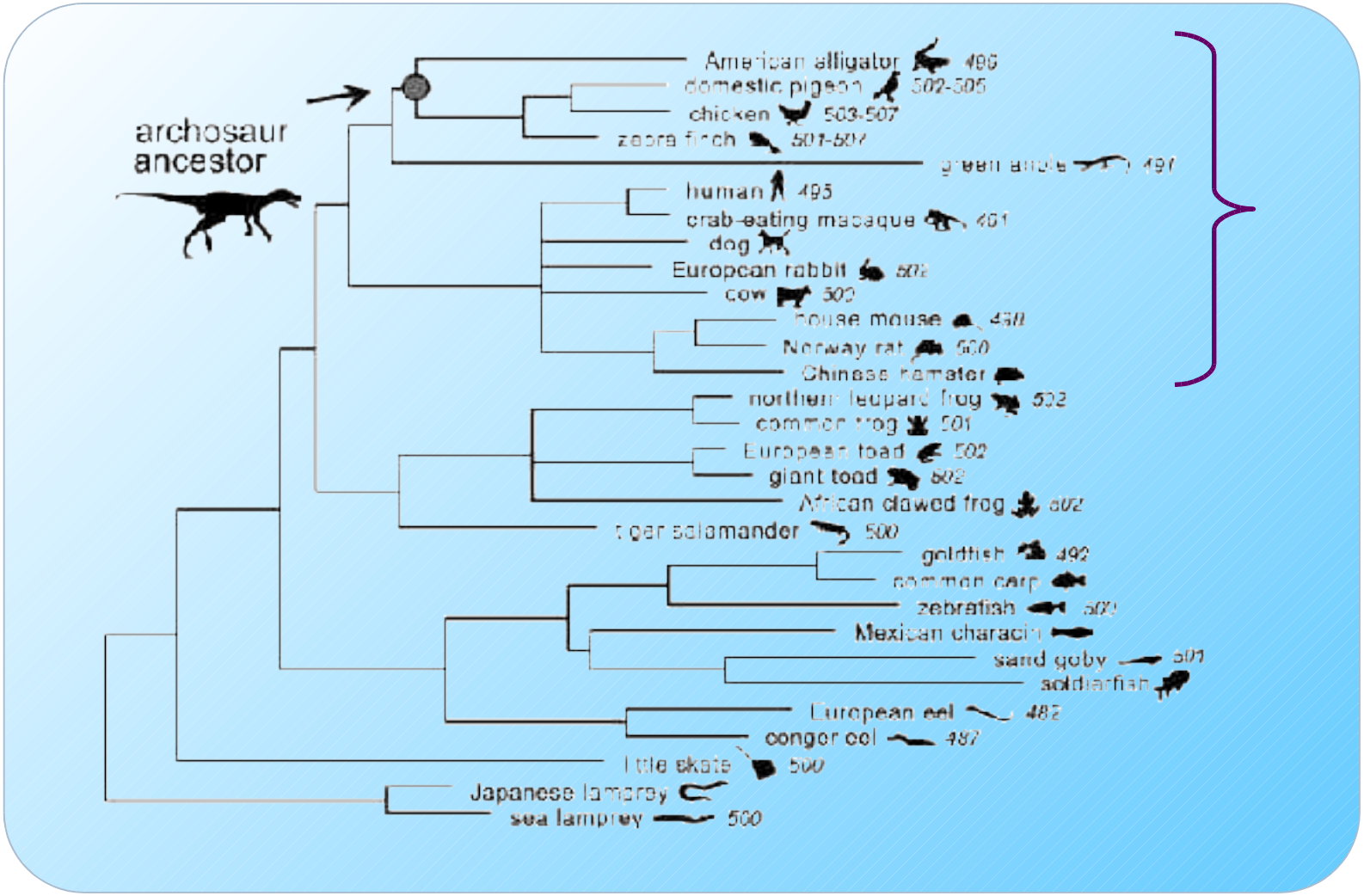
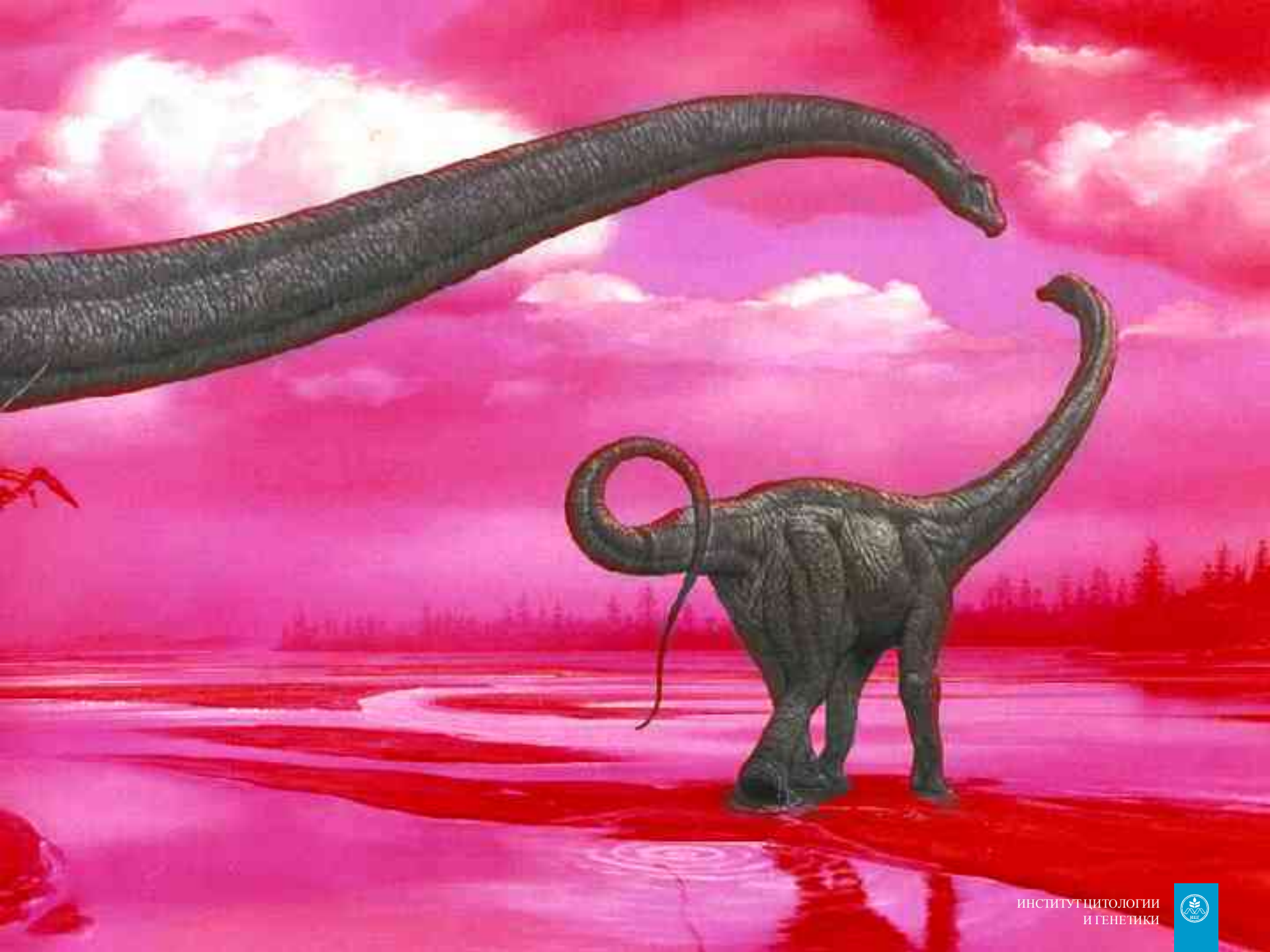



Figure is from Chang et al., *Mol. Biol. Evol.* 19(9):1483–1489. 2002