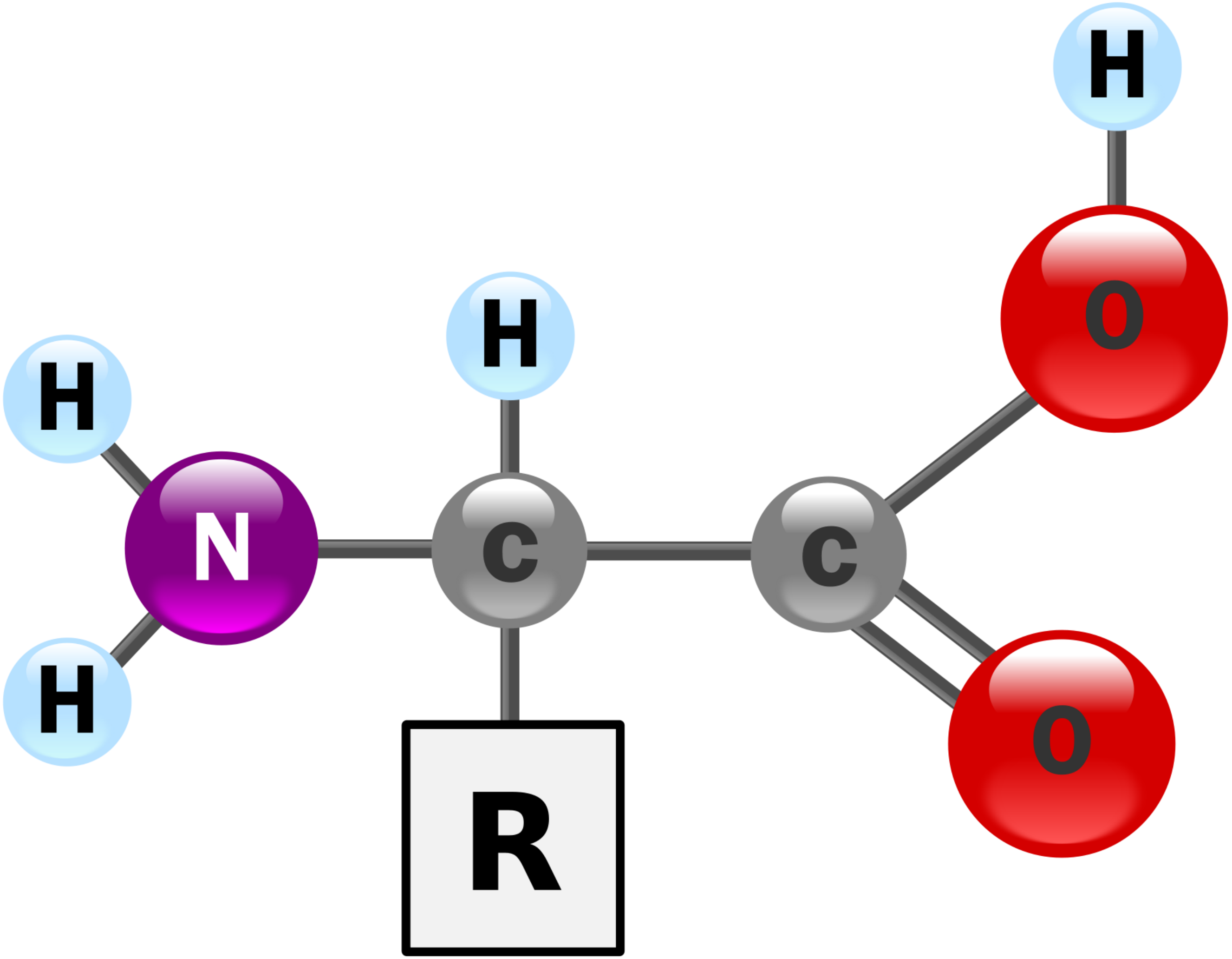
Метод анализа информационной структуры белков

Некрасов А.Н.

Международная конференция
«Проблема необратимости в классических
и квантовых динамических системах»
(Москва, 8–10 декабря 2011 г.)

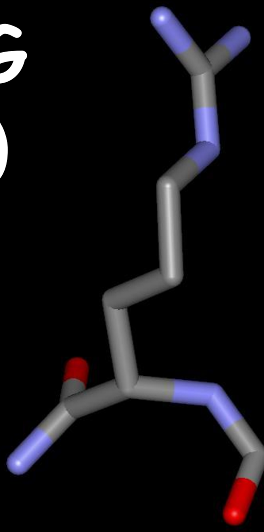
WHY PROTEINS ARE SO IMPORTANT?

PROTEINS ARE THE MAIN ACTOR IN LIVE SYSTEMS. PROTEINS PERFORM TRANSFORMATION OF ENERGY, SUBSTANCE, GENETIC INFORMATION AND FORM THE CONTROL SYSTEM.

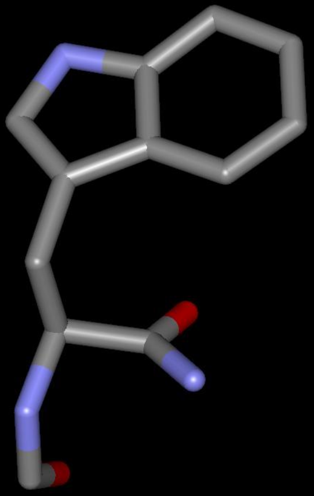

H
H
N
H
C
R
C
O
H
O

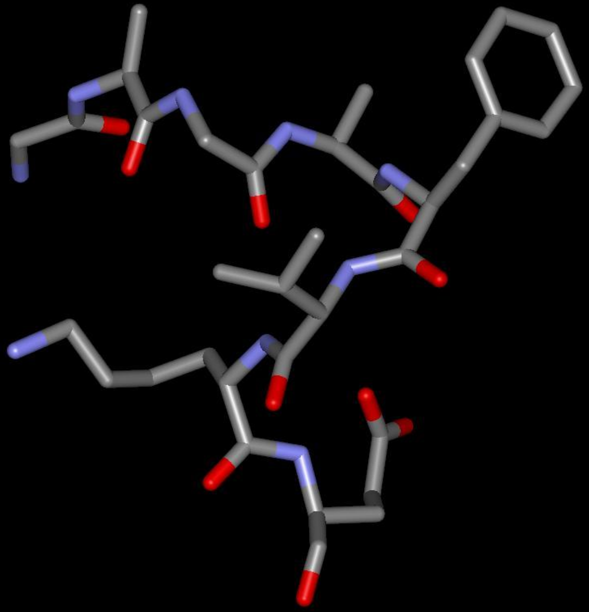
20 TYPES OF THE AMINO ACID RESIDUES (ACDEFGHIKLMNPQRSTVWY)

ARG
(R)

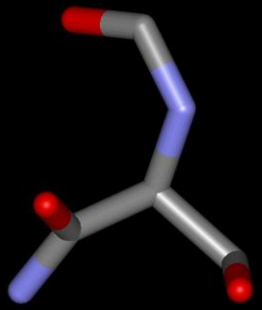


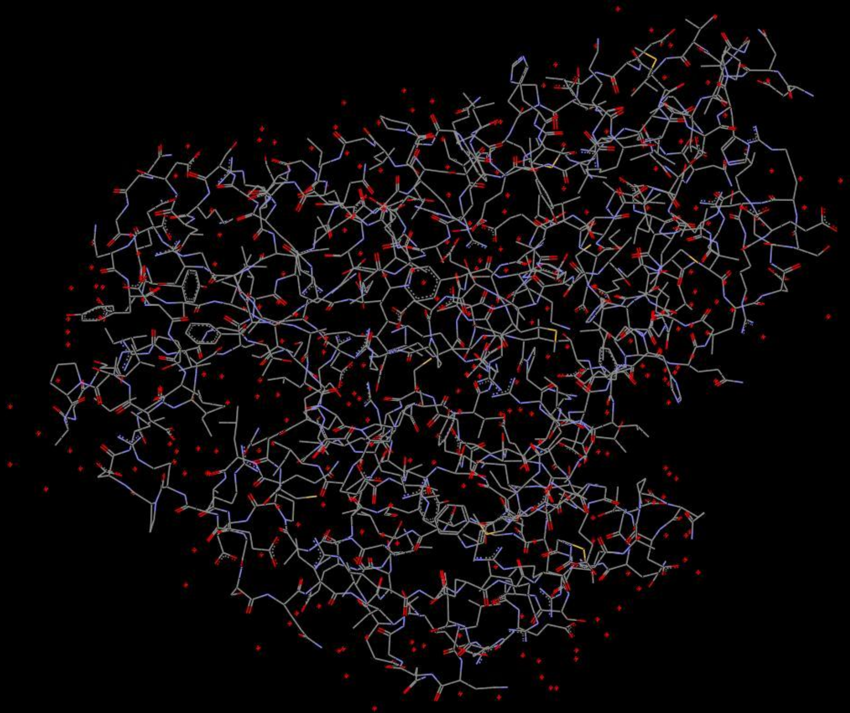
TRP
(W)



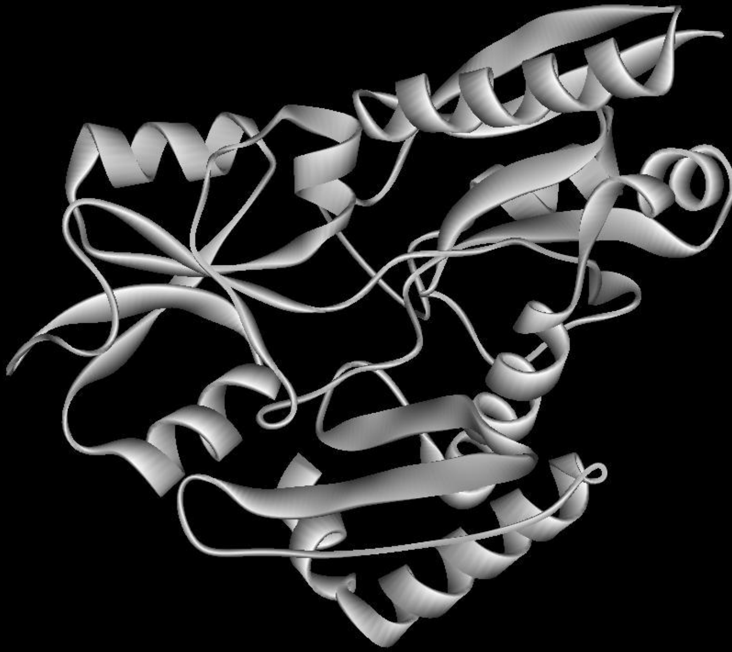


SER
(S)

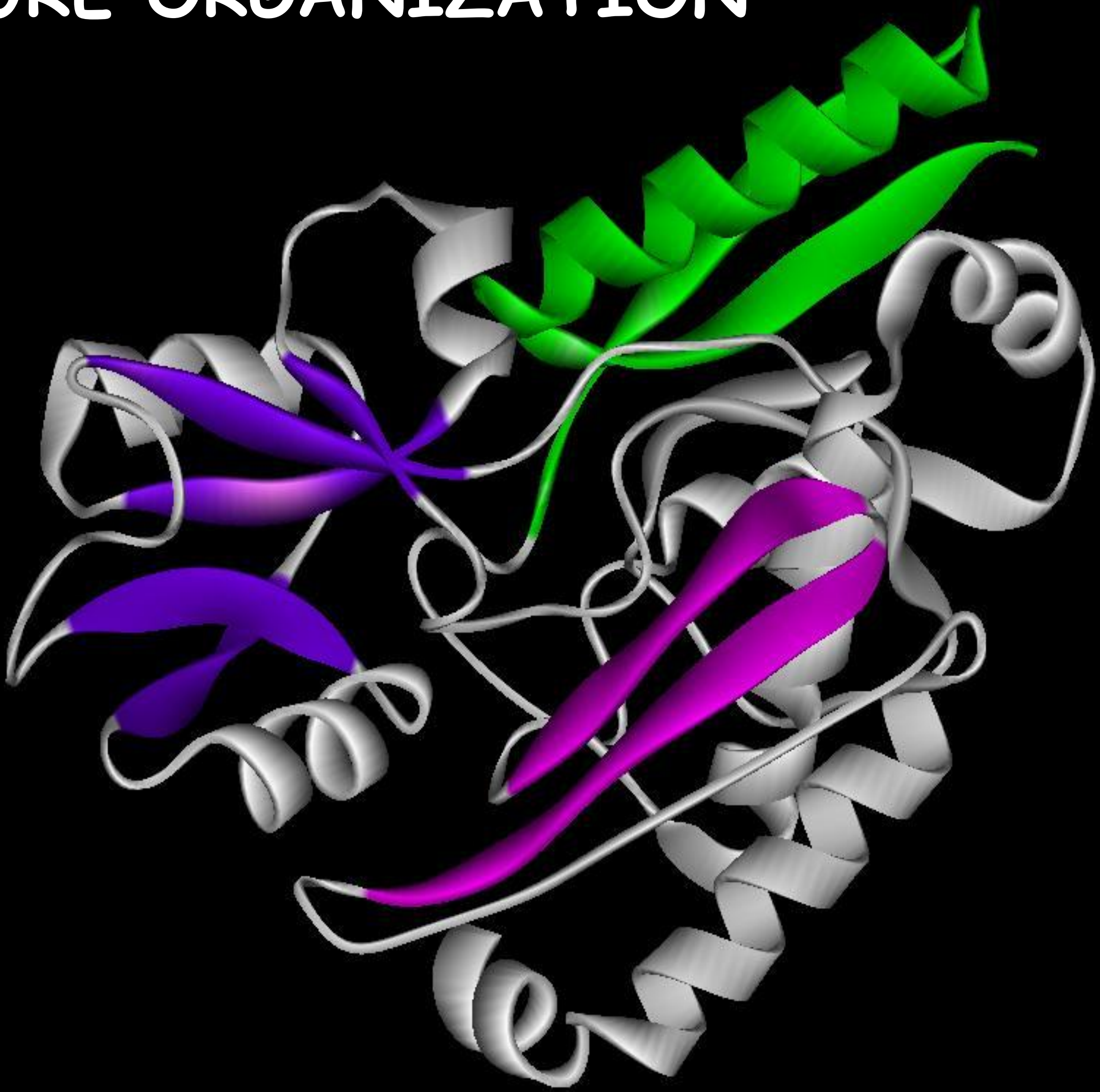




SPATIAL STRUCTURE OF PROTEIN



CLASSICAL VIEW ON THE PROTEIN SPATIAL STRUCTURE ORGANIZATION



SUPER-SECONDARY STRUCTURE

PROSITE DATABASE

PS00962 (411 264 000)

[LIVMFA]-x-{GPRV}-[LIVMFYC](2)-{LPC}-[STAC]-[GSTANQEKR]-[STALV]-[HY]-[LIVMF]-G

PS01328 (32 768 000 000 000)

[QR]-[IV]-x(4)-[TC]-D-x(2)-G-[IV]-V-x-[HF]-x(2)-[FY]

PS00606 (509 027 328 000 000)

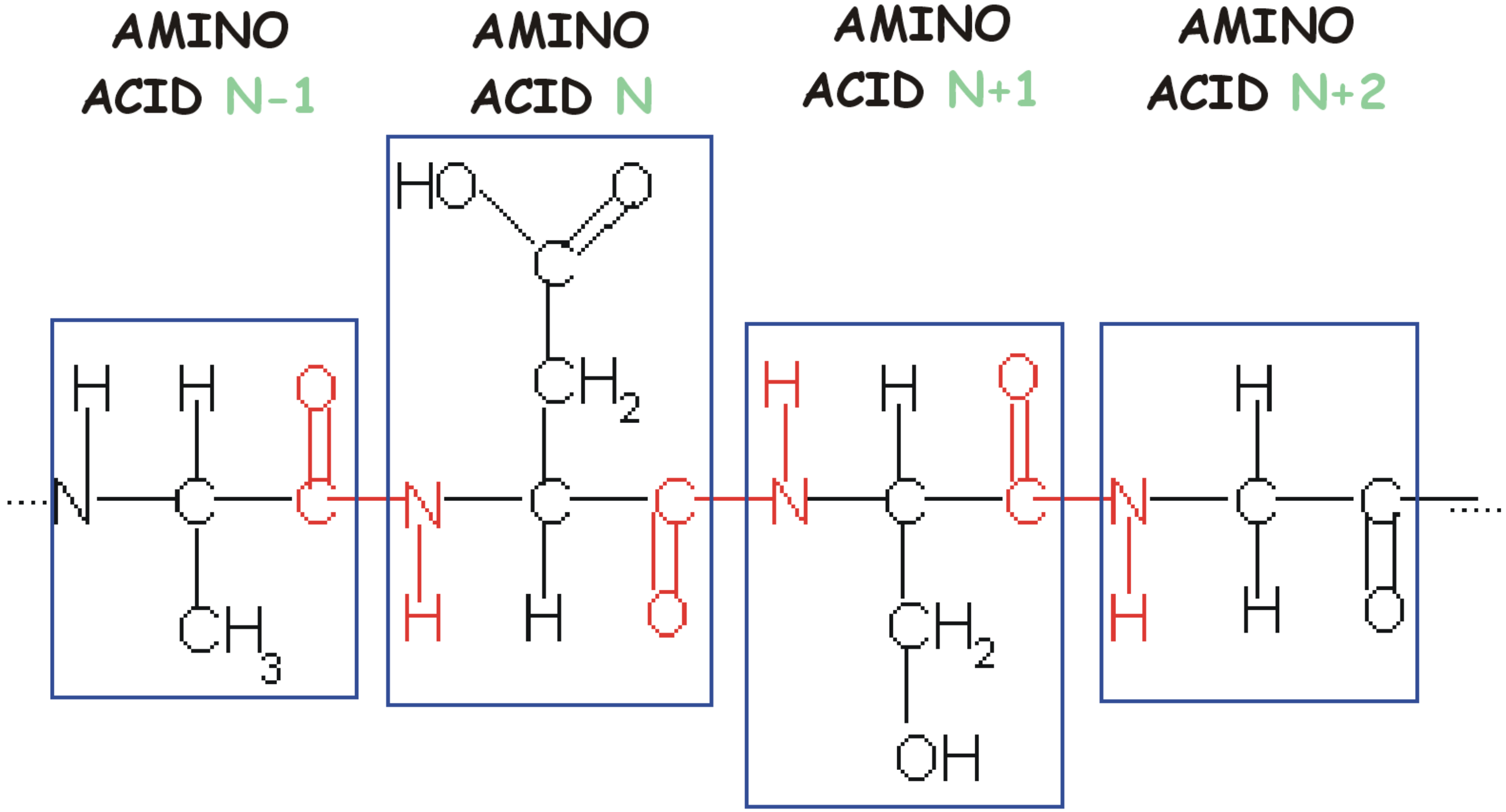
G-{A}-{KGR}-x(2)-[LIVMFTAP]-{R}-x-[AGC]-C-[STA](2)-[STAG]-x(2)-{LI}-[LIVMF]

Ptitsyn O.B., Volkenstein M.V.

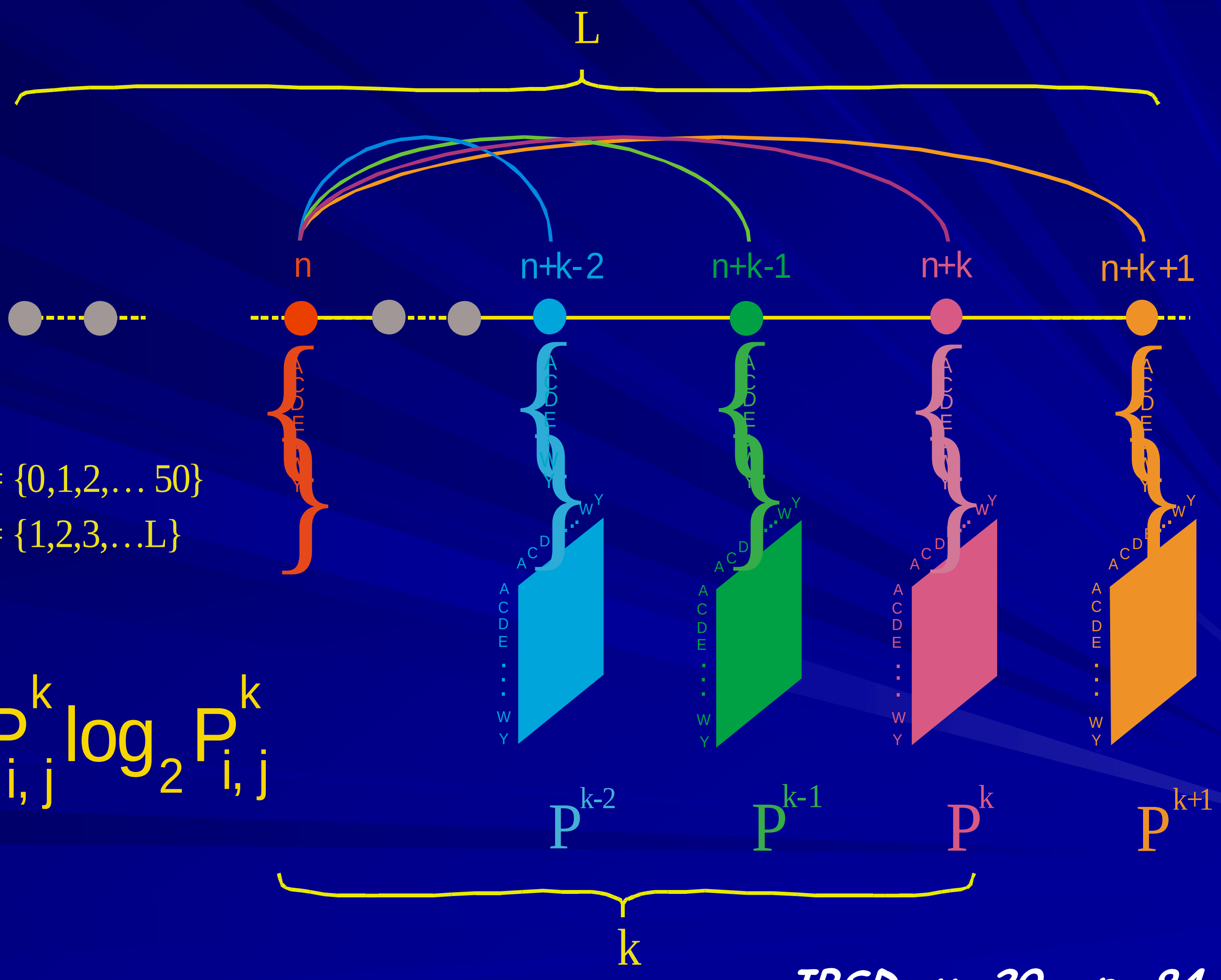
“Белки – слабо отредактированные гетерополимеры… ”

J. Biomol. Struct. Dyn. (1986) v.4(1), pp. 137-156

СТРУКТУРНЫЕ ЭЛЕМЕНТЫ БЕЛКОВ (КЛАССИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ)



METHOD OF PROTEIN SEQUENCE ANALYSIS



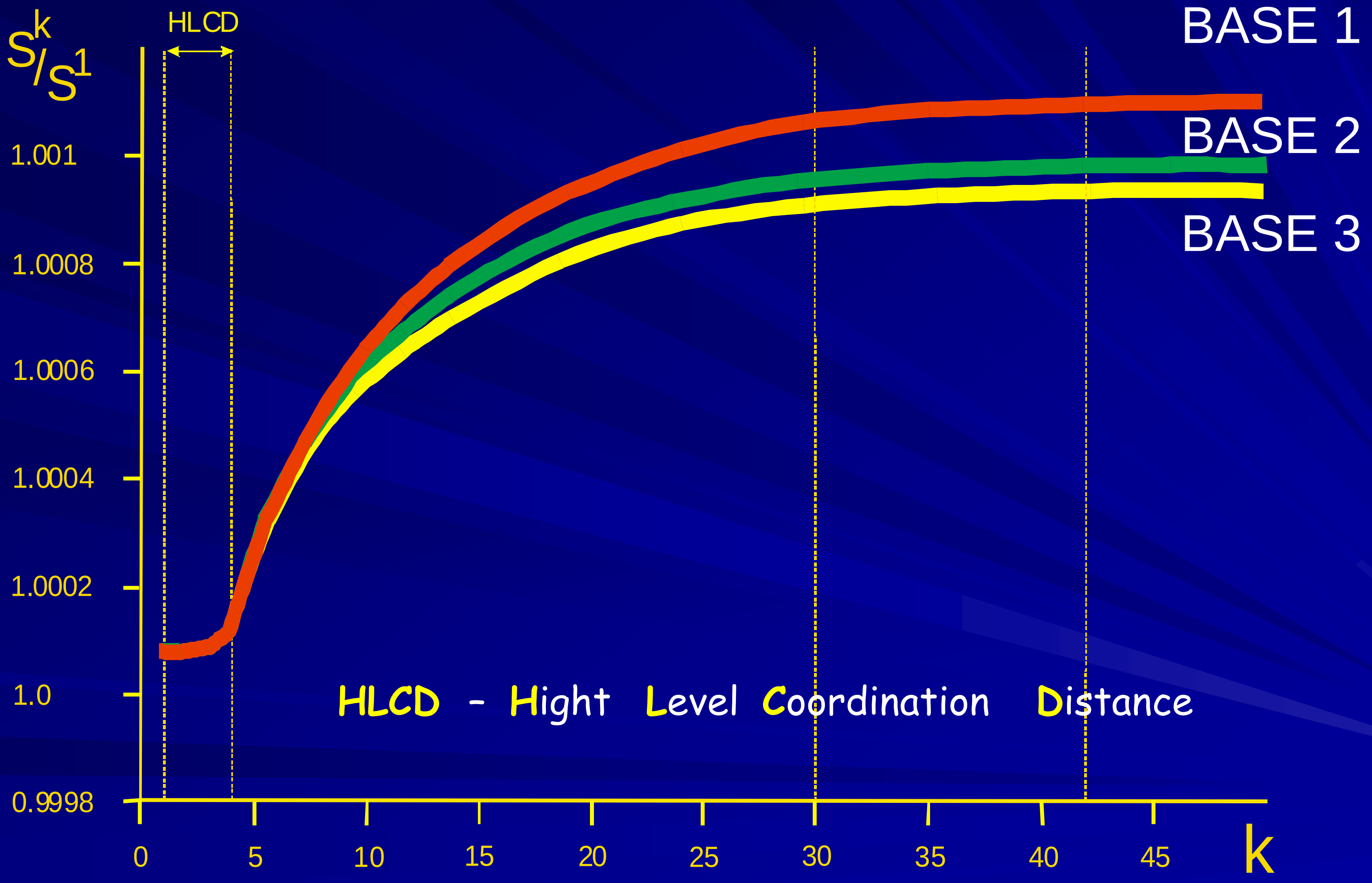
k = {0,1,2,… 50}

n = {1,2,3,…L}

$$S^k = -\sum_{i=1,20}^{20}\sum_{j=1,20}^{20} P_{i,j}^{k}\log_2 P_{i,j}^{k}$$

JBSD v. 20, p. 84

POSITION SHANNON'S ENTROPY



HLCD – **H**ight **L**evel **C**oordination **D**istance

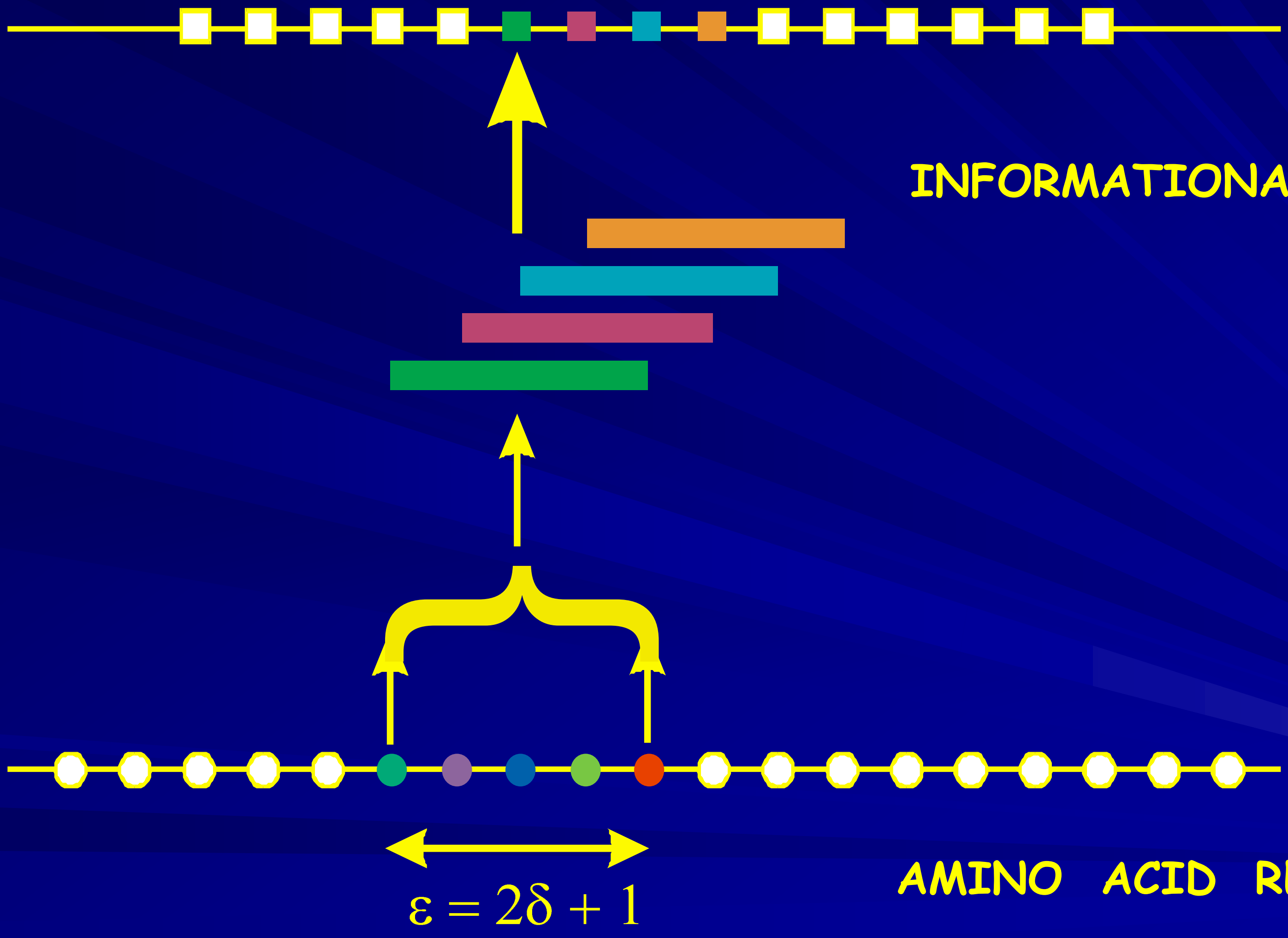
JBSD v. 20, p. 84

THE OBTAINED DATA SHOW:

- EXISTANCE OF LONG AREAS IN PROTEIN SEQUENCES WITH HIGHER LEVEL CORRELATIONS BETWEEN AMINO ACID RESIDUES

- ON SHORT DISTANCES BETWEEN THE RESIDUES HIGH AND CONSTANT LEVEL OF CORRELATION BETWEEN THE RESDUES (LOW LEVEL OF SHANNON'S ENTROPY) IS OBSERVED. THIS FACT ALLOWS US CONSIDER SHORT FRAGMENTS OF PROTEINS SEQUENCE AS A BASIC ELEMENTS OF PROTEIN SEQUENCE – "INFORMATIONAL UNITS".

REPRESENTATION OF SEQUENCE AS A SET OF INFORMATIONAL UNITS



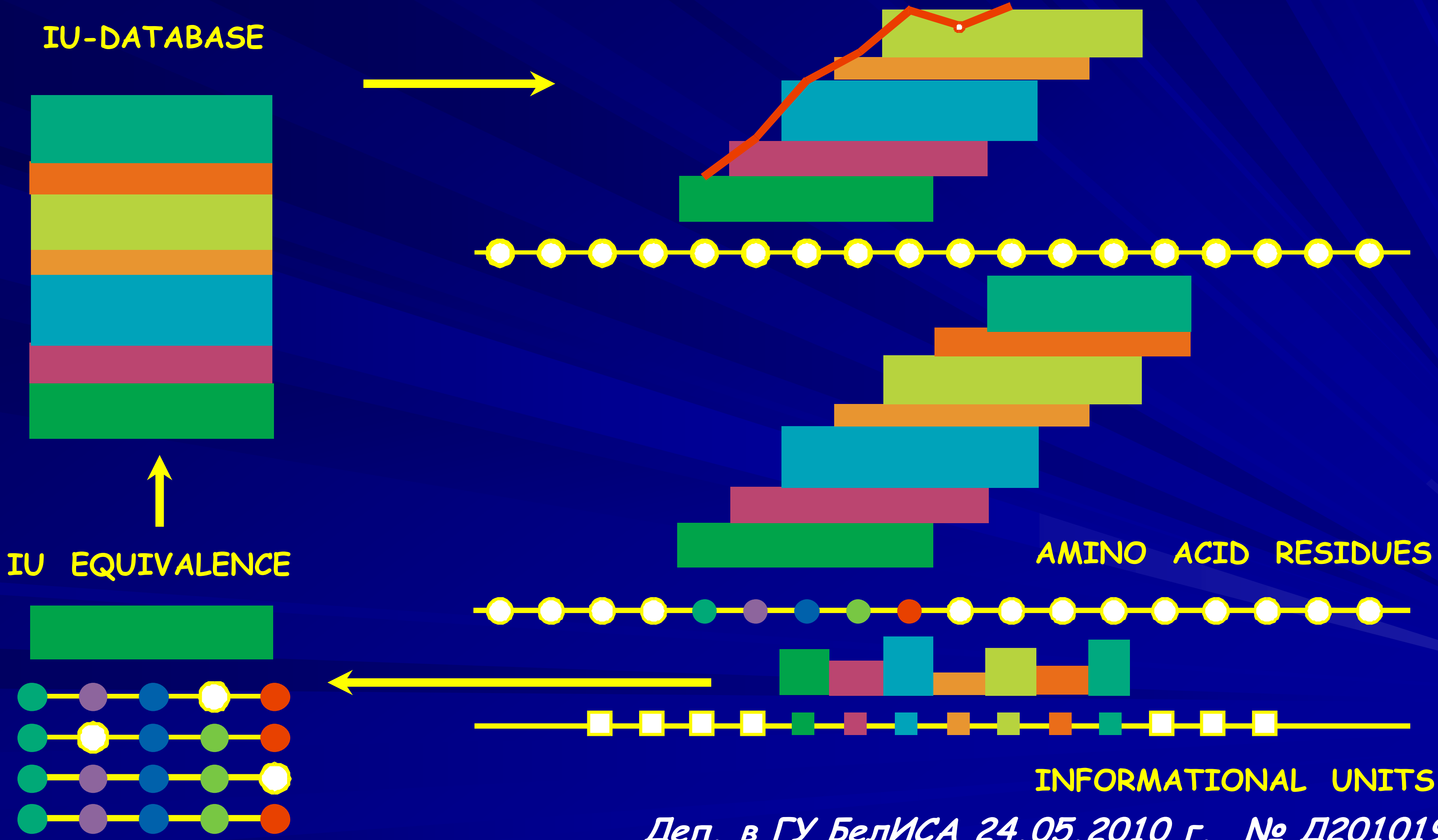
INFORMATIONAL UNITS

$\varepsilon = 2\delta + 1$

AMINO ACID RESIDUES

Деп. в ГУ БелИСА 24.05.2010 г., № Д201019

COMPUTATION OF INFORMATIONAL UNITS POPULATION PROFILE OF PROTEINS



IU-DATABASE

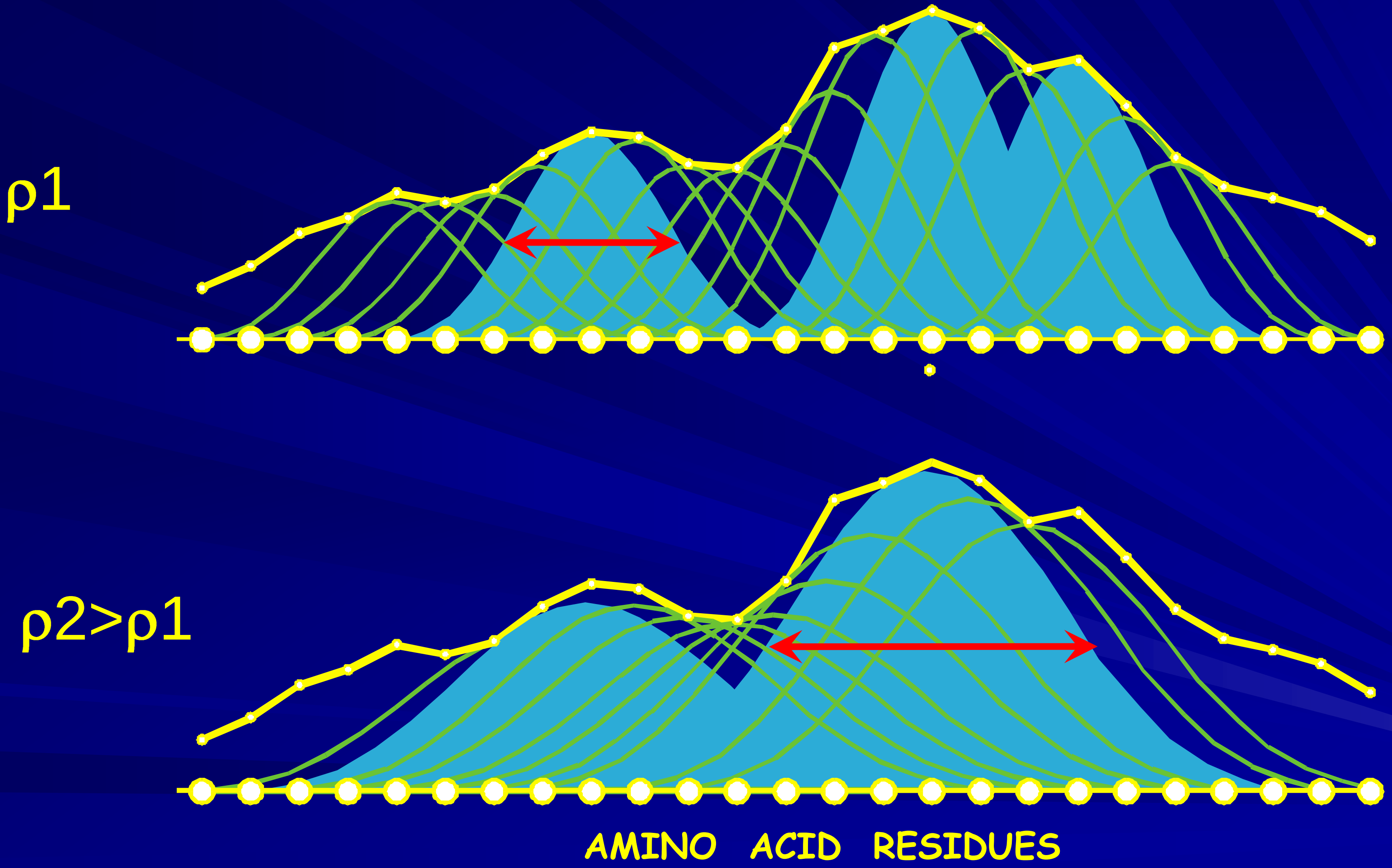
IU EQUIVALENCE

AMINO ACID RESIDUES

INFORMATIONAL UNITS

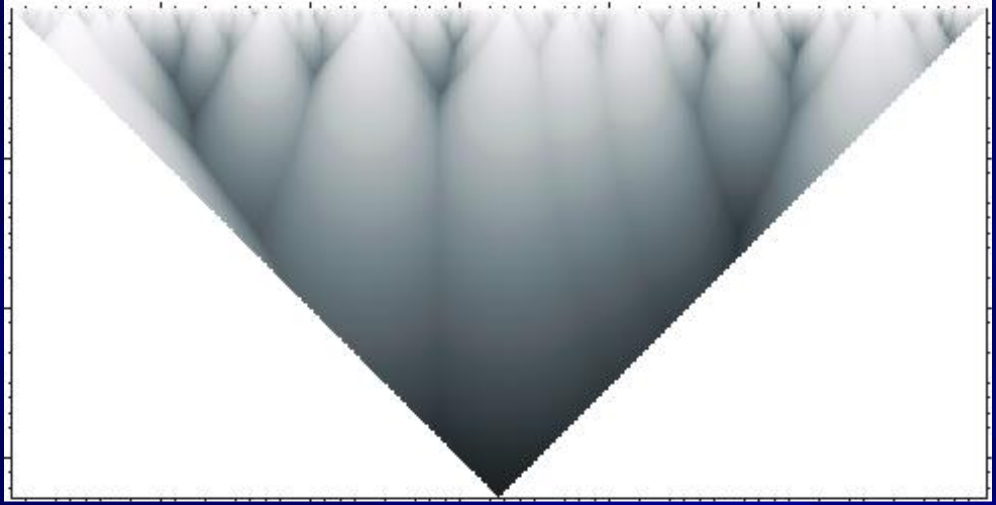
Деп. в ГУ БелИСА 24.05.2010 г., № Д201019

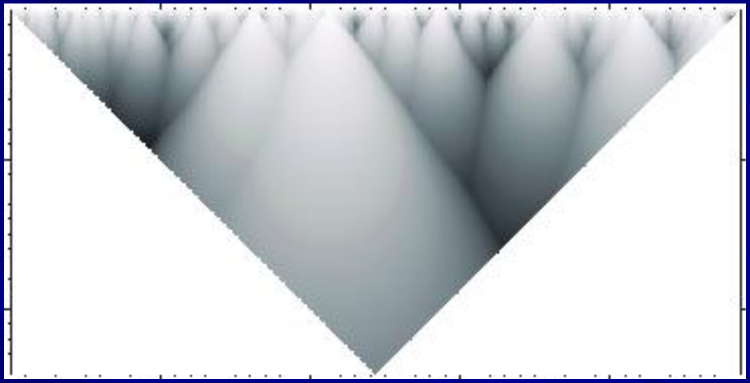
GAUSSIAN APPROXIMATION OF POPULATION PROFILE OF PROTEINS

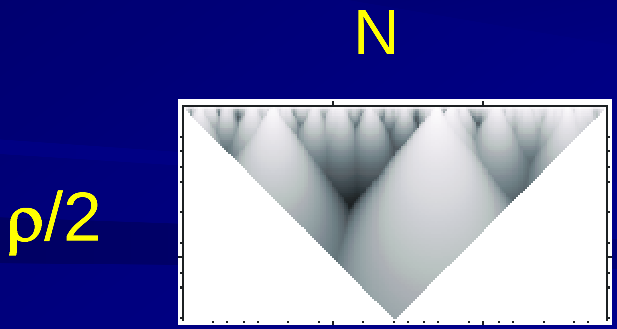


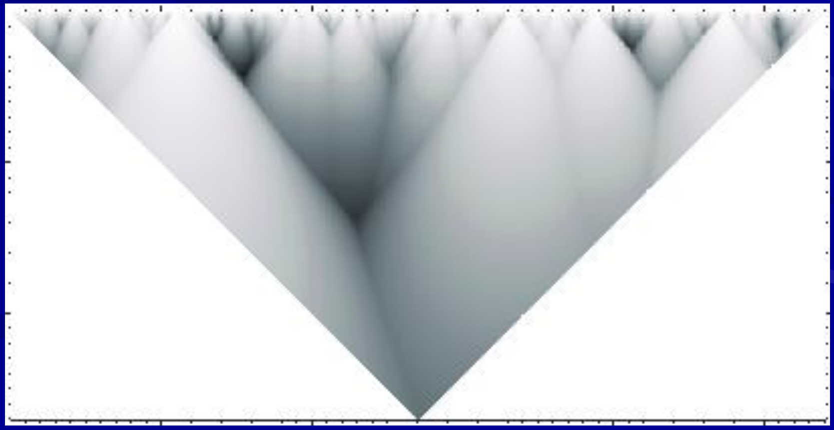
Деп. в ГУ БелИСА 24.05.2010 г., № Д201019

GRAPHICAL REPRESENTATION OF PROTEIN INFORMATIONAL STRUCTURE





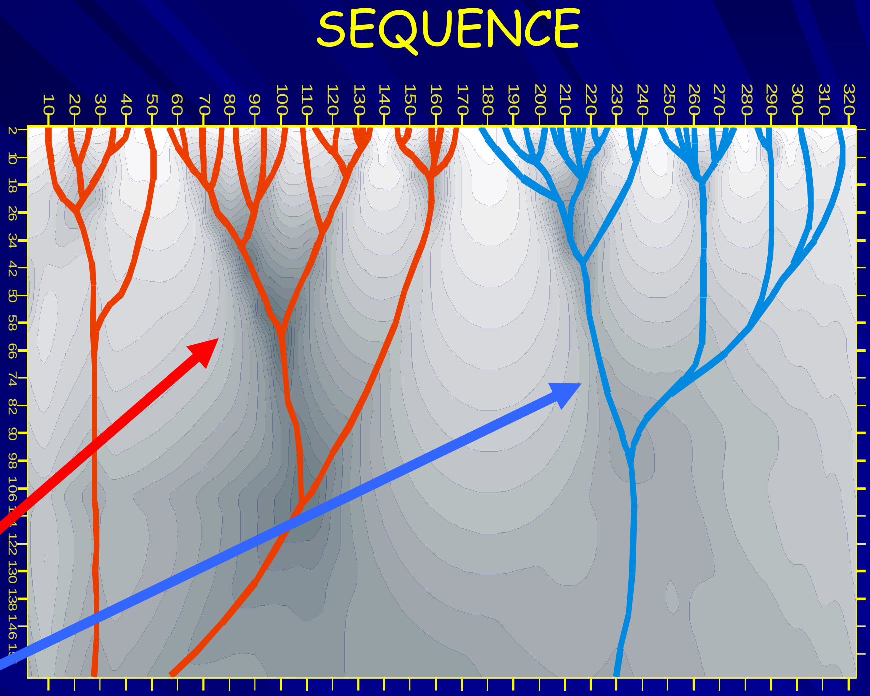




Деп. в ГУ БелИСА 24.05.2010 г., № Д201019

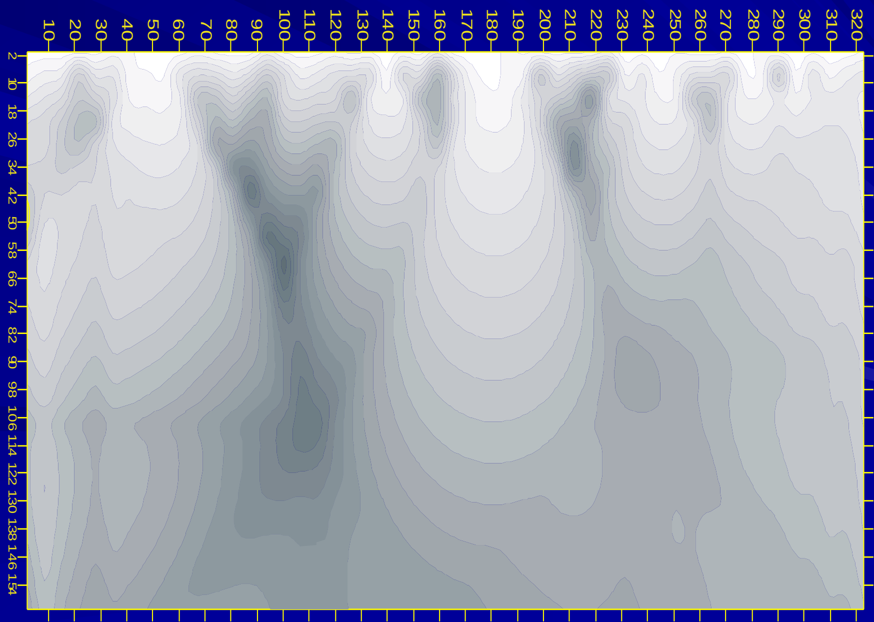
PEPSIN (1AM5.PDB)

Graphical representation of protein **Informational Structure**



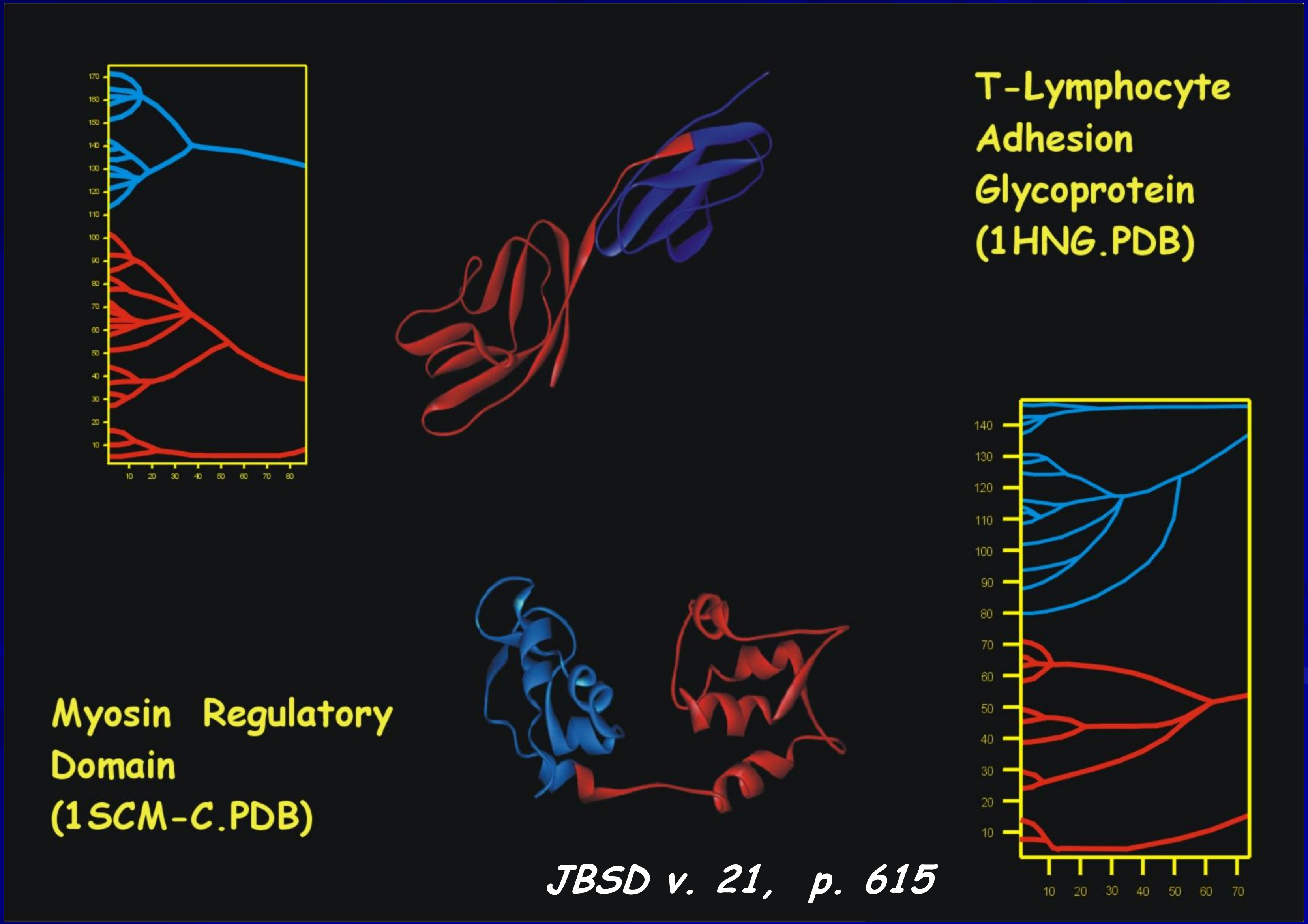
$\rho/2$

ELIS - High Rank **EL**ements of **I**nformational **S**tructure



JBSD v. 21, p. 615

HIGH RANK ELIS RELATIONS WITH RESPECT TO STRUCTURAL DOMAINS



T-Lymphocyte Adhesion Glycoprotein (1HNG.PDB)

Myosin Regulatory Domain (1SCM-C.PDB)

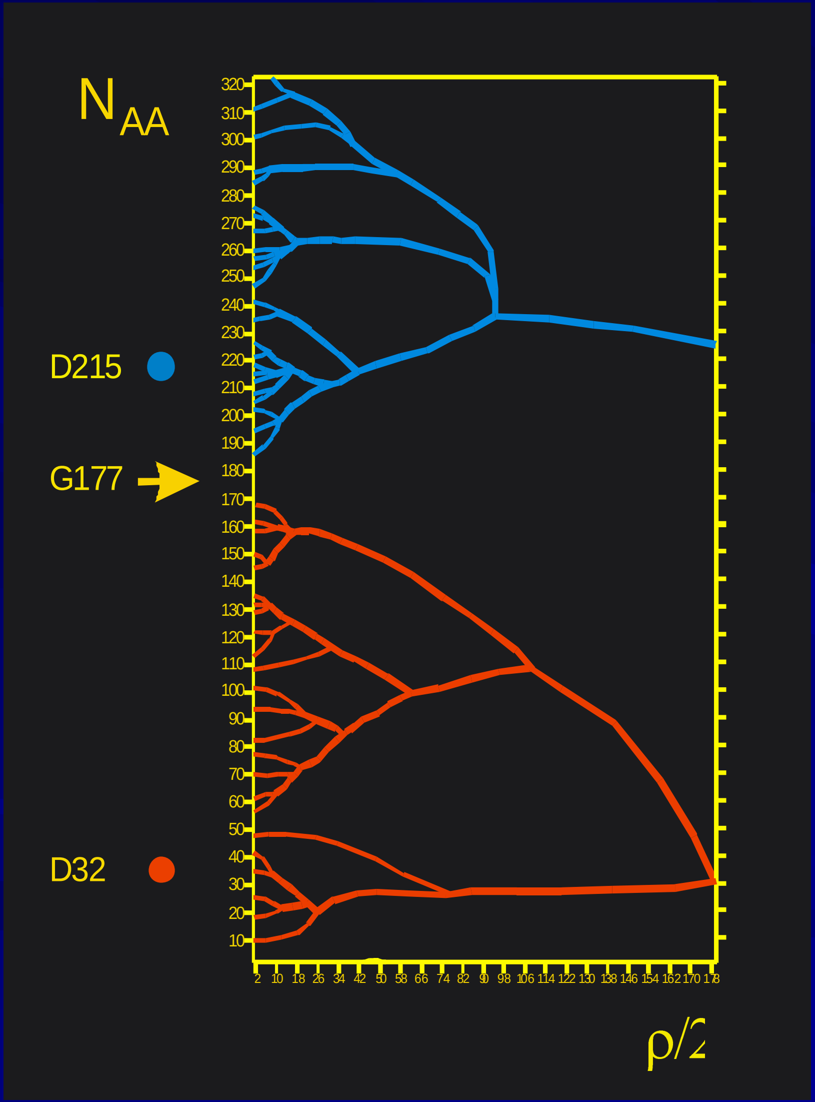
JBSD v. 21, p. 615

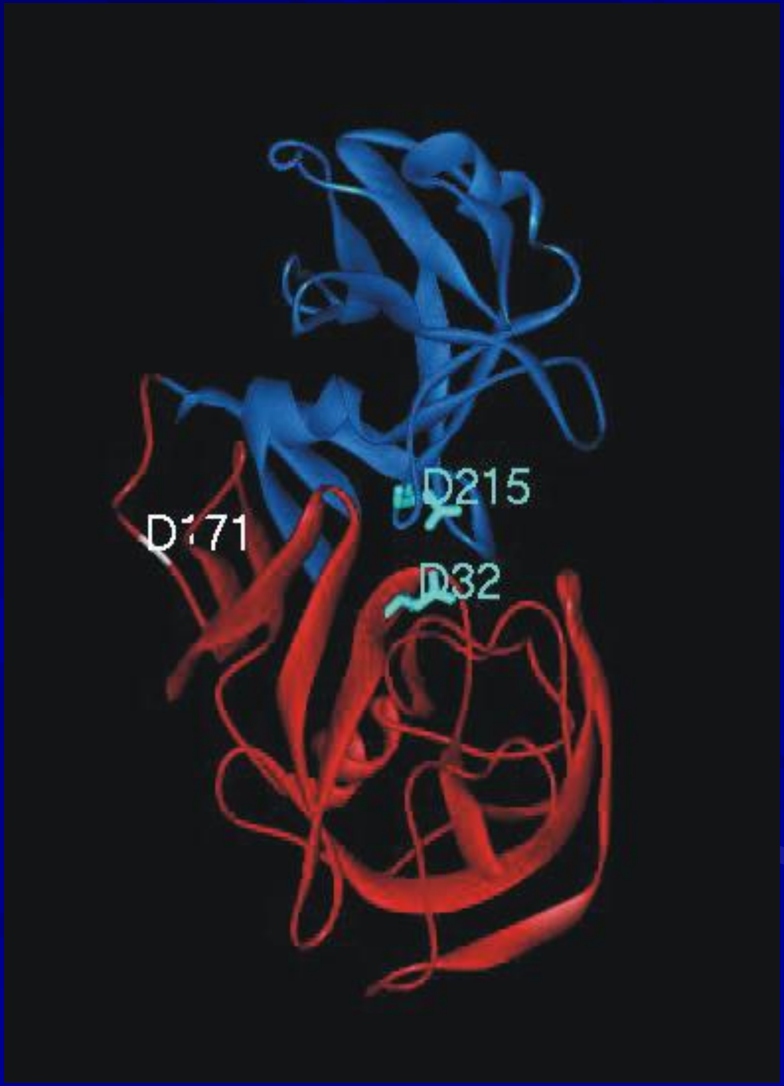
CATALITIC REACTIONS EQUATION
(Michaelis - Menten Equation)

$$E + S \overset{k_S}{\rightleftharpoons} ES^* \longrightarrow EP^* \xrightarrow{k_{cat}} E + P$$

E – ENZYME
S – SUBSTRATE
P - PRODUCT

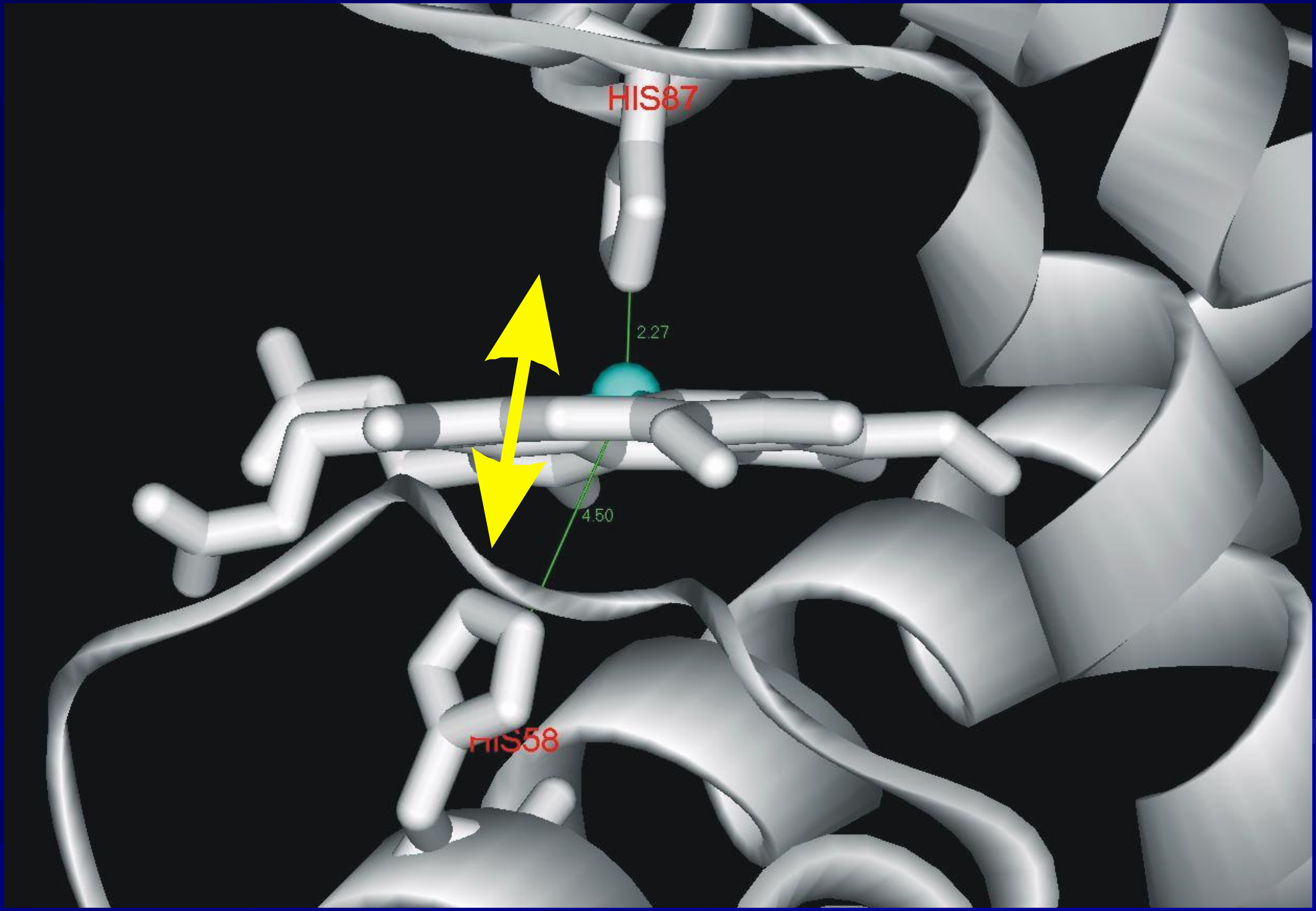
PEPSIN FROM GADUS MORHUA
(EC: 3.4.23.1) 1AM5.PDB





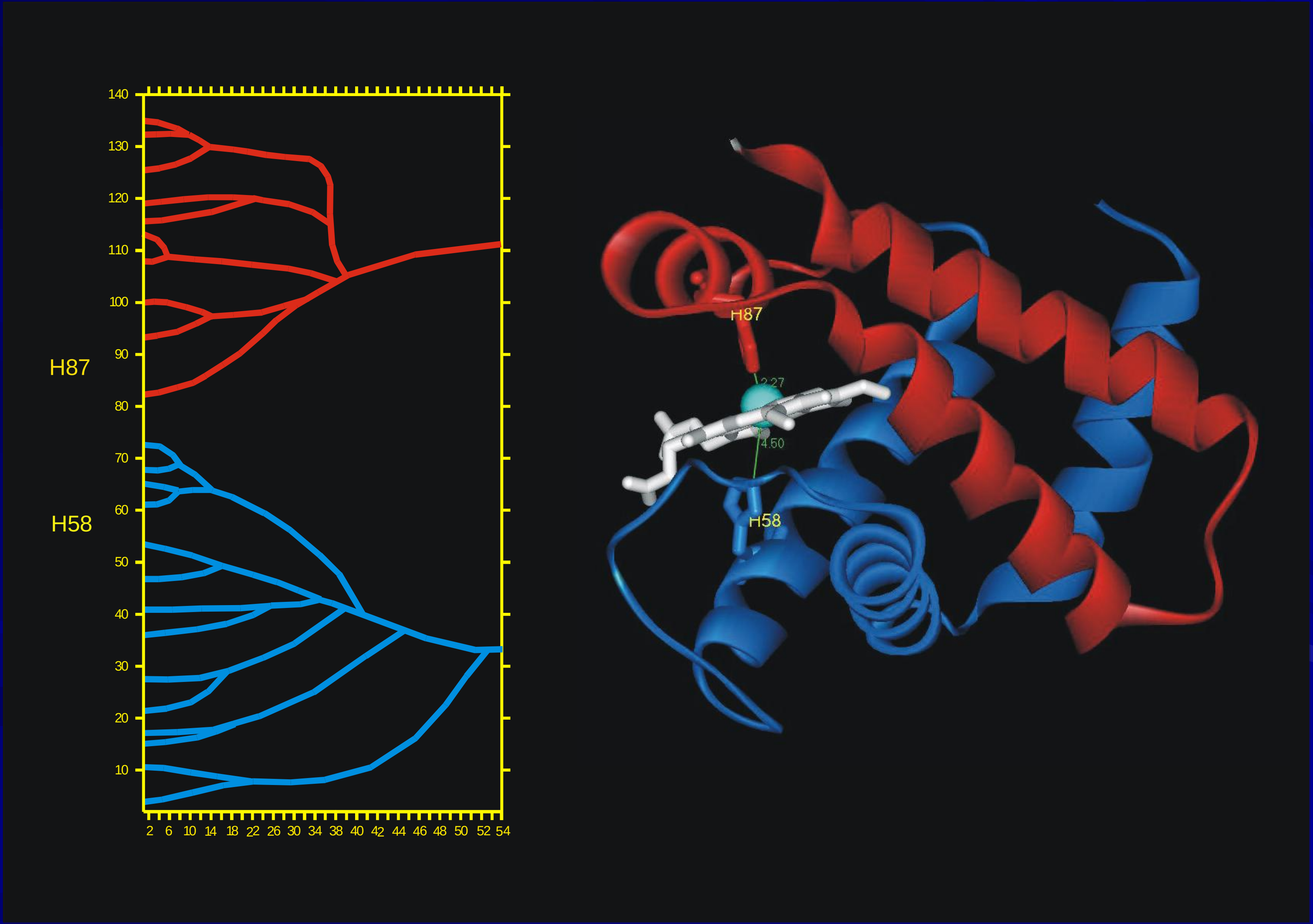
JBSD (2008) v. 25, p. 554

ACTIVE CENTRE OF HEMOGLOBIN β-CHAIN

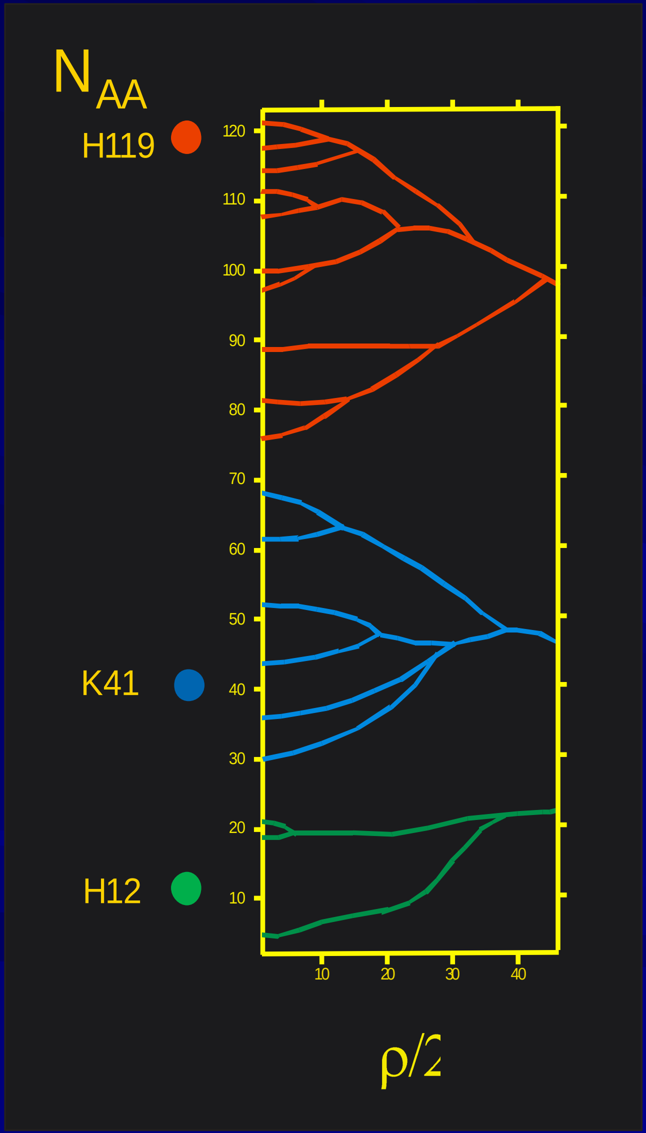


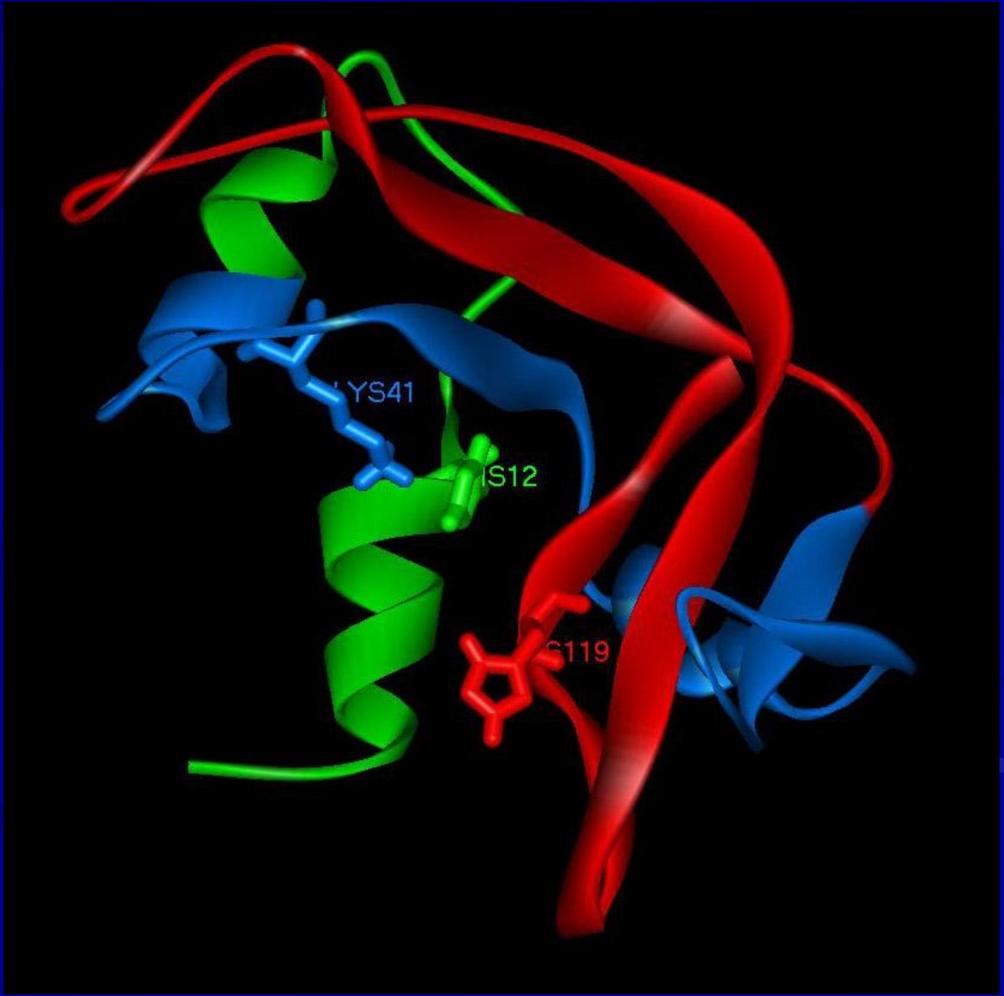
Paoli et. al. JMB (1996) v. 256, p.775

HIGH RANK ELIS OF HEMOGLOBIN β-CHAIN (1BZ0.PDB)



RNAse A FROM BOS TAURUS (EC: 3.1.27.5) 1AQP.PDB

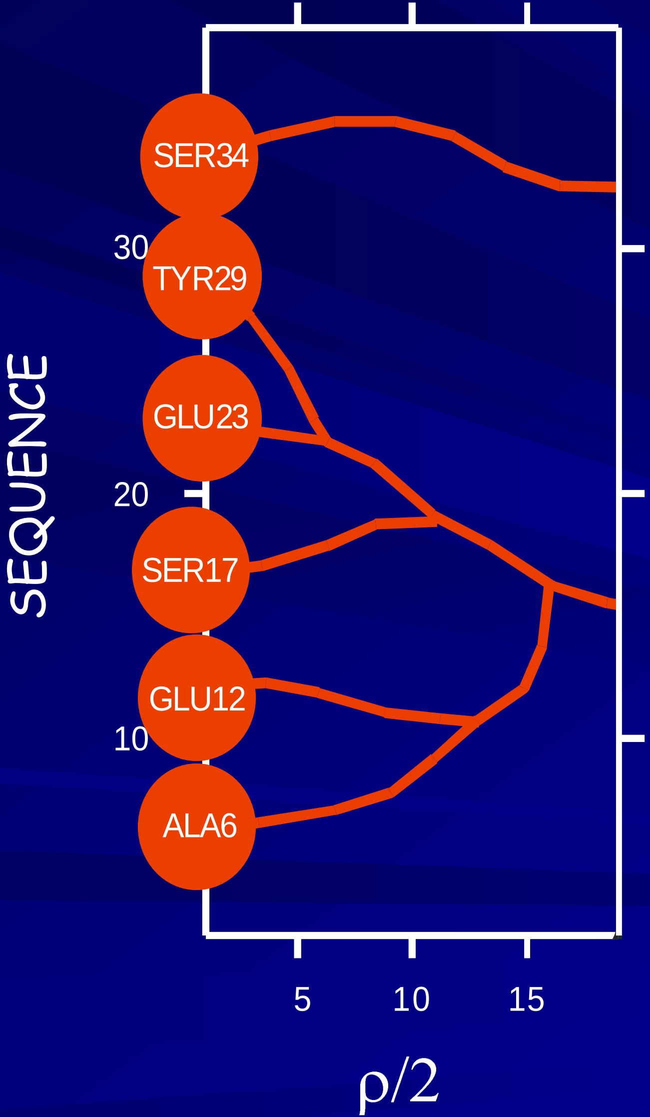


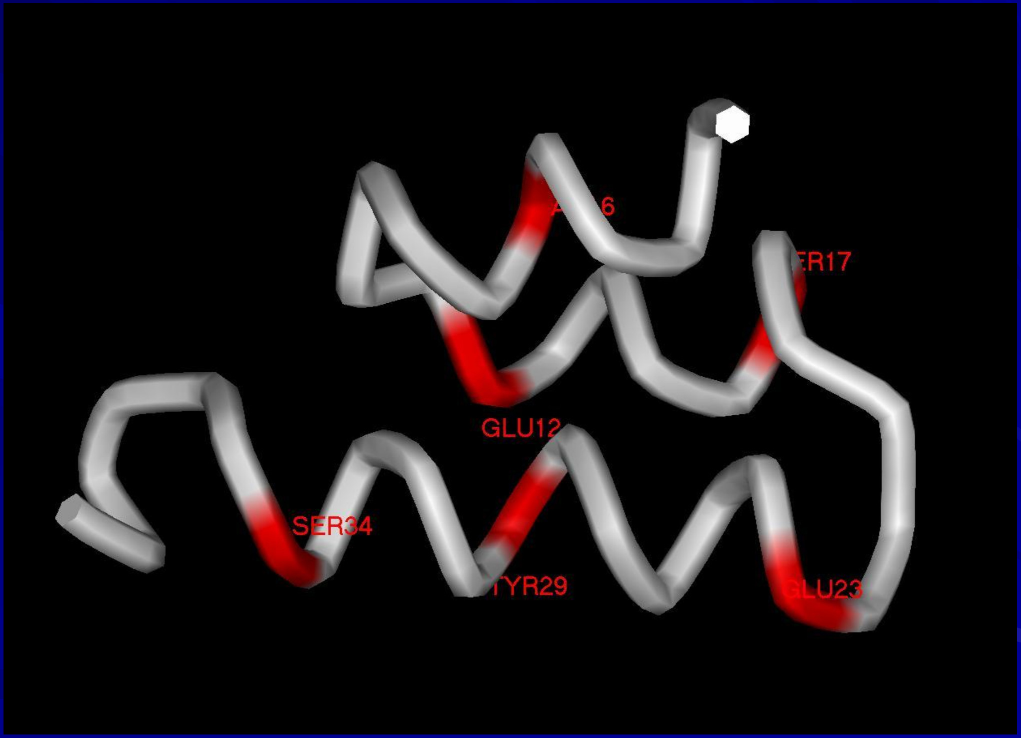


JBSD (2008) v. 25, p. 554

PHEROMONE ER-1

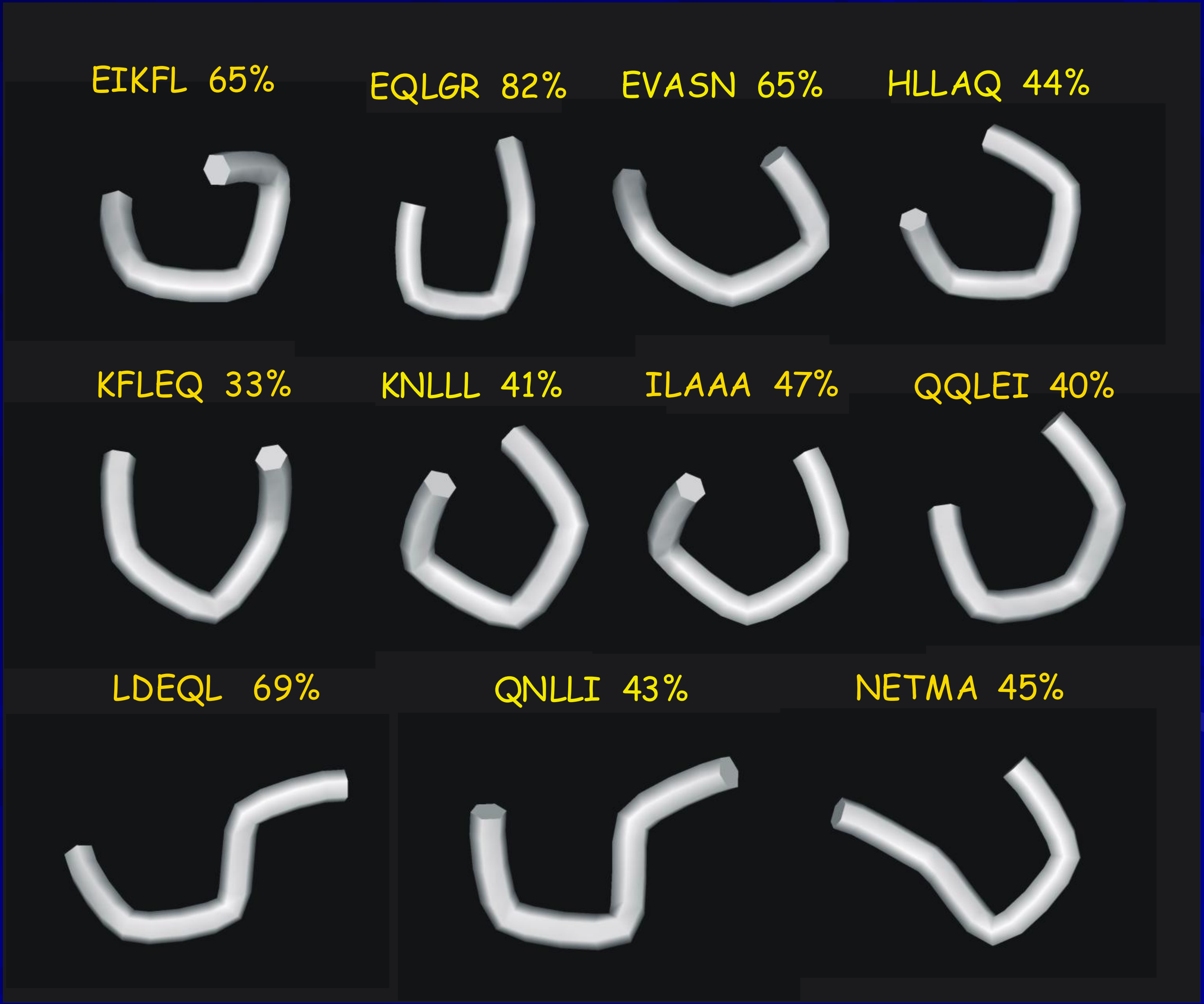
(2ERD.PDB)



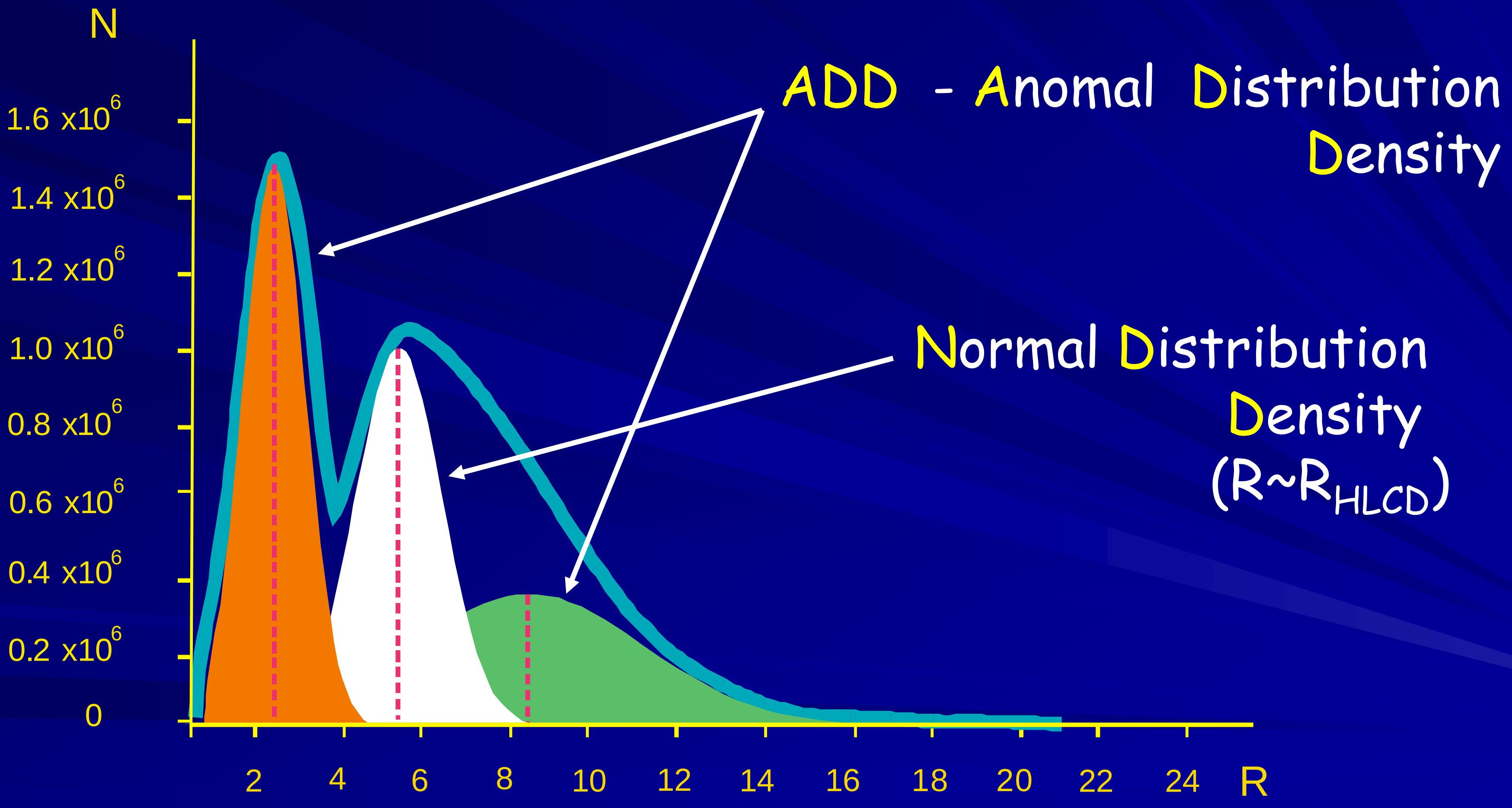


J Biomol Struct Dyn. vol. 28(1), p. 85.

CONFORMATIONS OF LARGEST CLUSTERS OF FIRST RANK ELIS AFTER MD



DISTRIBUTION OF FIRST RANK ELIS DISTANCE

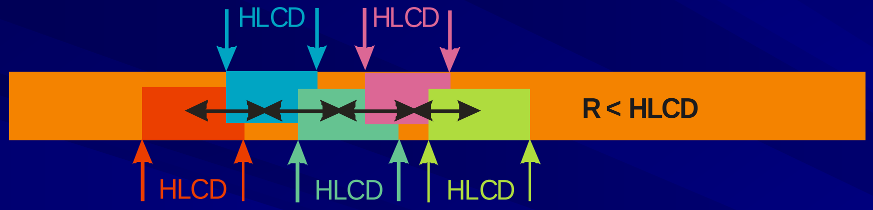


ADD - Anomal Distribution Density

Normal Distribution Density ($R \sim R_{HLCD}$)

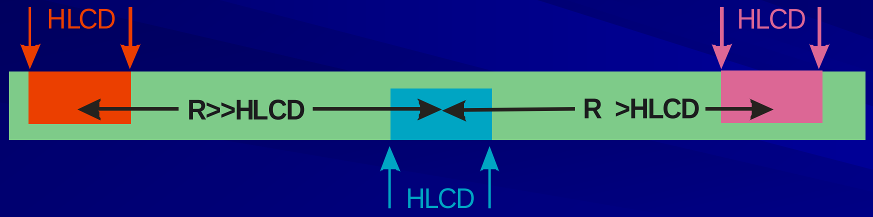
J Biomol Struct Dyn. vol. 28(1), p. 85.

PROTEIN SITES CLASSIFICATION BY DENSITY OF FIRST RANK ELIS & TOPOLOGICAL PROPETIES OF SITES



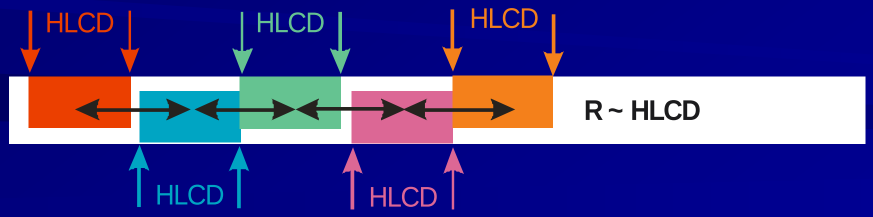
ADD+

TOPOLOGICALLY
CONSTRAINED SITE



ADD-

TOPOLOGICALLY
VARIABLE SITE



NORMAL

TOPOLOGICALLY
STANDARD SITE

J Biomol Struct Dyn. vol. 28(1), p. 85.

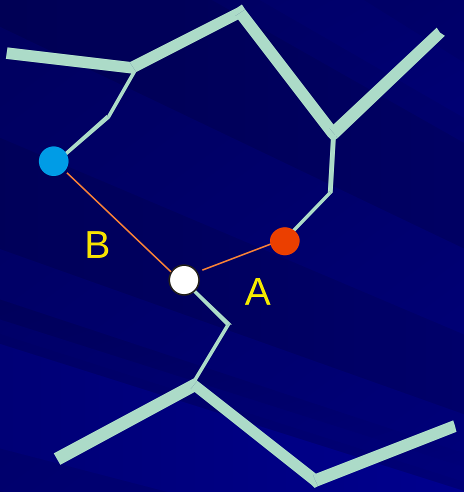
FACTORS THAT INFLUENCE ON INTERACTION BETWEEN POLYPEPTIDE CHAINS

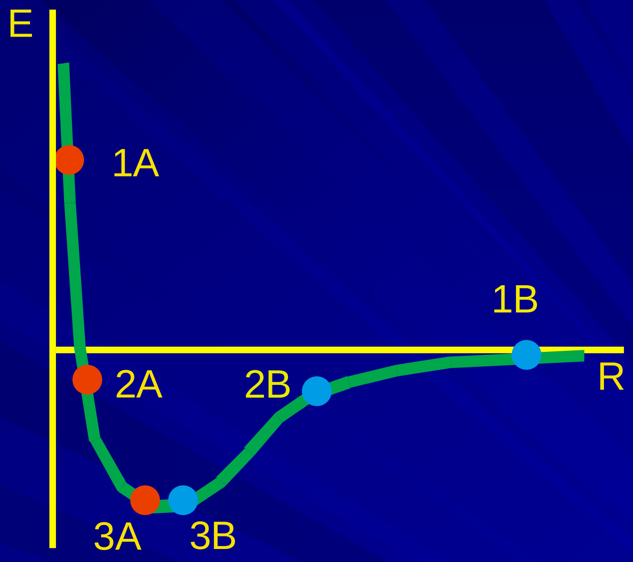
- EXISTENCE OF FUNCTIONAL GROUPS OF AMINO ACID RESIDUES THAT PROVIDE INTERACTION

- ABILITY OF POLYPEPTIDE CHAIN TO CHANGE CONFORMATION IN ORDER TO PROVIDE OPTIMUM DISTANCE FOR INTERACTING FUNCTIONAL GROUPS

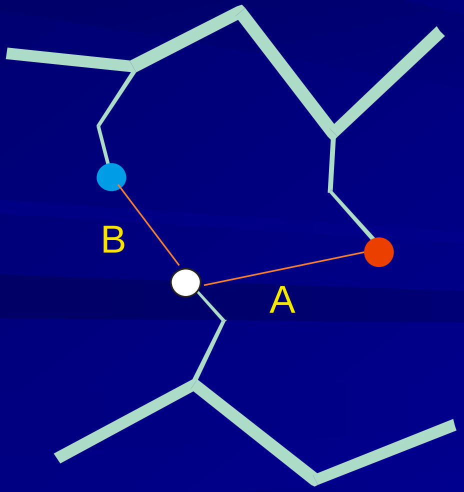
ADAPTABLE CONFORMATIONAL CHANGES WHILE POLYPEPTIDE CHAINS INTERACTION

1

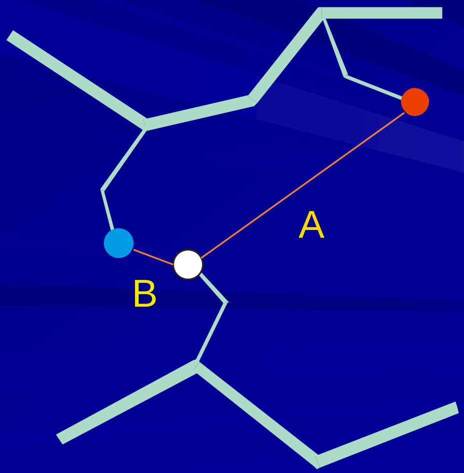




2

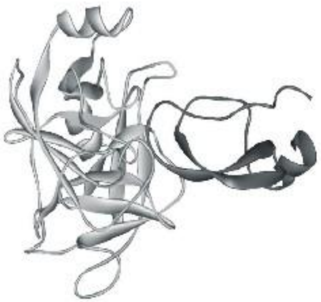


3

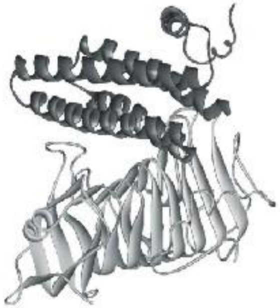


SPATIAL STRUCTURES OF ENZYME-INHIBITOR COMPLEXES

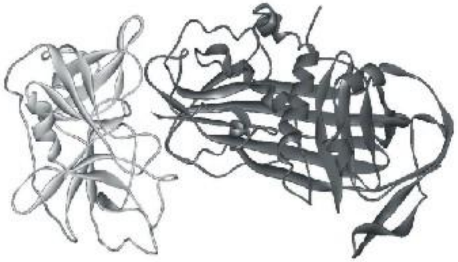
TRSP/BPTI
(3TGI.PDB)



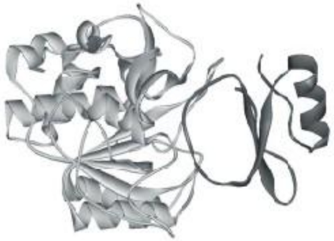
PME 1/PMEI
(1XG2.PDB)



PPE/α_1-PI
(2D26.PDB)

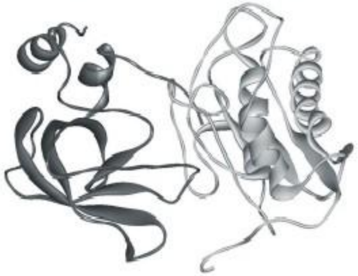


SUBT/CI-2A
(1A10.PDB)



RNase A/RI
(1DFJ.PDB)

MMP-3/TIMP-1
(1OO0.PDB)



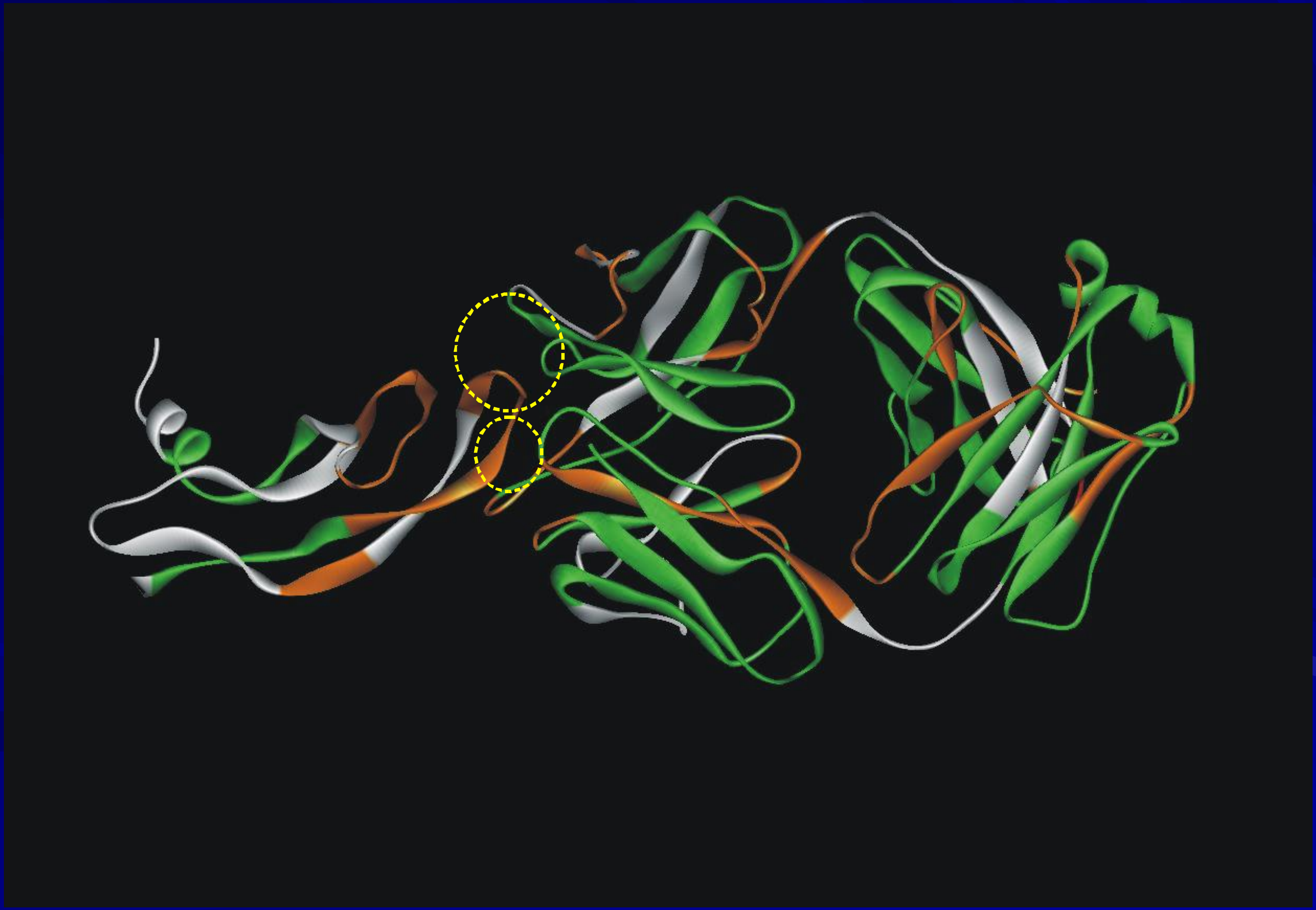
J Biomol Struct Dyn. vol. 28(1), p. 85.

IS TYPE CONTACTS REALIZED IN HYDROLASE-INGIBITOR COMPLEXES

TYPE CONTACT	TRPS/BPTI		SUBT/CI2A		PME/PMEI		RNAse/RI		PPE/aPI		MMP3/TIMP		
ADD-/ADD+	93.6	95.2	39.0	72.7	33.8	85.9	31.4	96.1	14.3	92.9	9.6	62.8	84.2
ADD-/NORM	0.0		33.7		31.0		23.5		14.3		46.9		
ADD-/ADD-	1.6		0.0		21.1		41.2		64.3		6.3		
NORM/NORM	0.0	4.8	11.7	27.3	5.6	14.1	0.0	3.9	0.0	7.1	24.6	37.2	15.8
ADD+/NORM	3.2		13.0		8.5		3.9		7.1		12.6		
ADD+/ADD+	1.6		2.6		0.0		0.0		0.0		0.0		

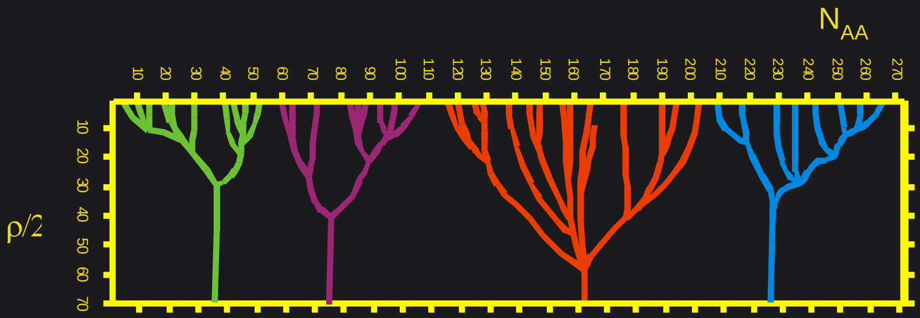
J Biomol Struct Dyn. vol. 28(1), p. 85.

VASCULAR ENDOTHELIAL GROWTH FACTOR IN COMPLEX WITH A NEUTRALIZING ANTIBODY (1BJ1.PDB)

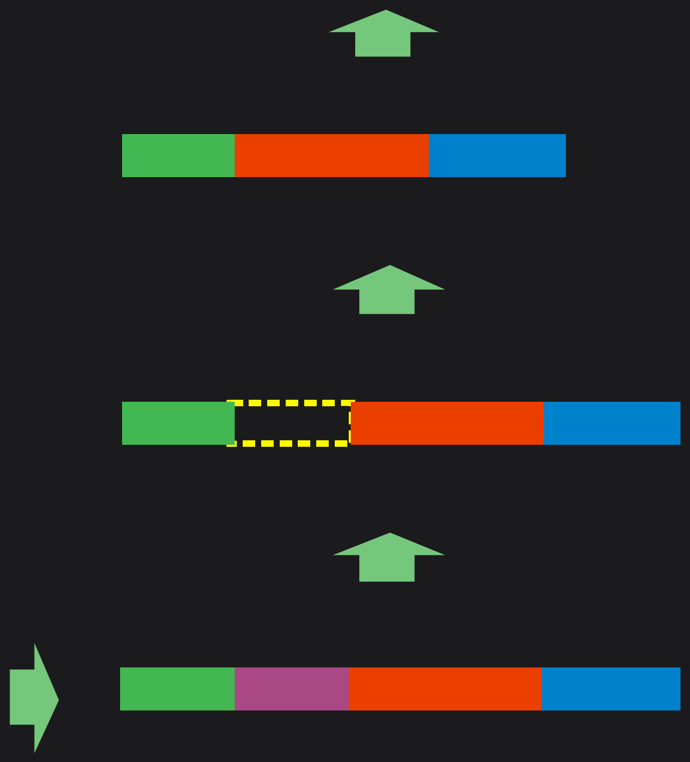


METHOD OF NEW RECOMBINANT PROTEIN DESIGN

```
>PROTEIN  SEQUENCE
AQTVPYGIPL IKADKVQAQG FKGANVKVAV LDTGIQASHP DLNVVGGASF
VAGEAYNTDG NGHGTHVAGT VAALDNTTGV LGVAPSVSLY AVKVLNSSGS
GSYSGIVSGI EWATTNGMDV INMSLGGASG STAMKQAVDN AYARGVVVVA
AAGNSGNSGS TNTIGYPAKY DSVIAVGAVD SNSNRASFSS VGAELEVMAP
GAGVYSTYPT NTYATLNGTS MASPHVAGAA ALILSKHPNL SASQVRNRLS
STATYLGSSF YYGKGLINVE AAAQ*
```



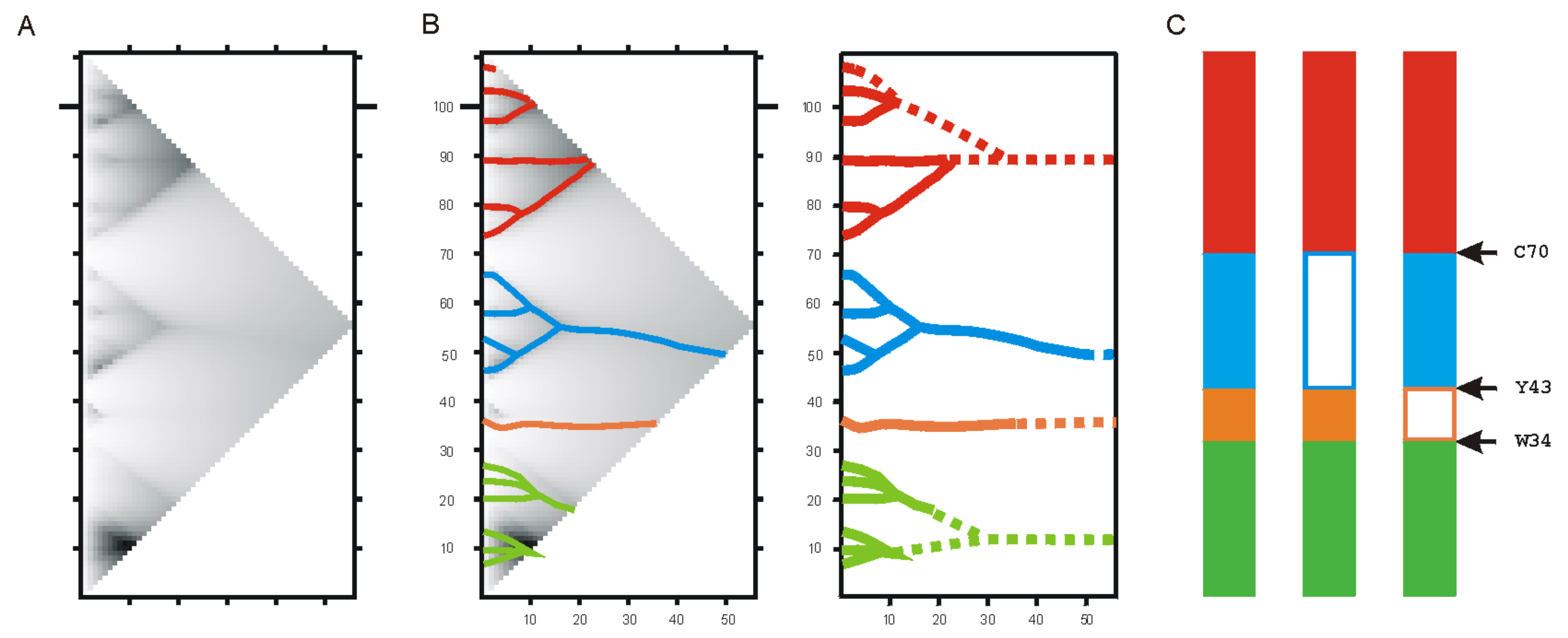
```
>PROTEIN  SEQUENCE
AQTVPYGIPL IKADKVQAQG FKGANVKVAV LDTGIQASHP DLNVVGGASF
VAGEAYNTDG NGHGTHVAGT VAALDNTTGV LGVAPSVSLY AVKVLNSSGS
GSYSGIVSGI EWATTNGMDV INMSLGGASG STAMKQAVDN AYARGVVVVA
AAGNSGNSGS TNTIGYPAKY DSVIAVGAVD SNSNRASFSS VGAELEVMAP
GAGVYSTYPT NTYATLNGTS MASPHVAGAA ALILSKHPNL SASQVRNRLS
STATYLGSSF YYGKGLINVE AAAQ*
```



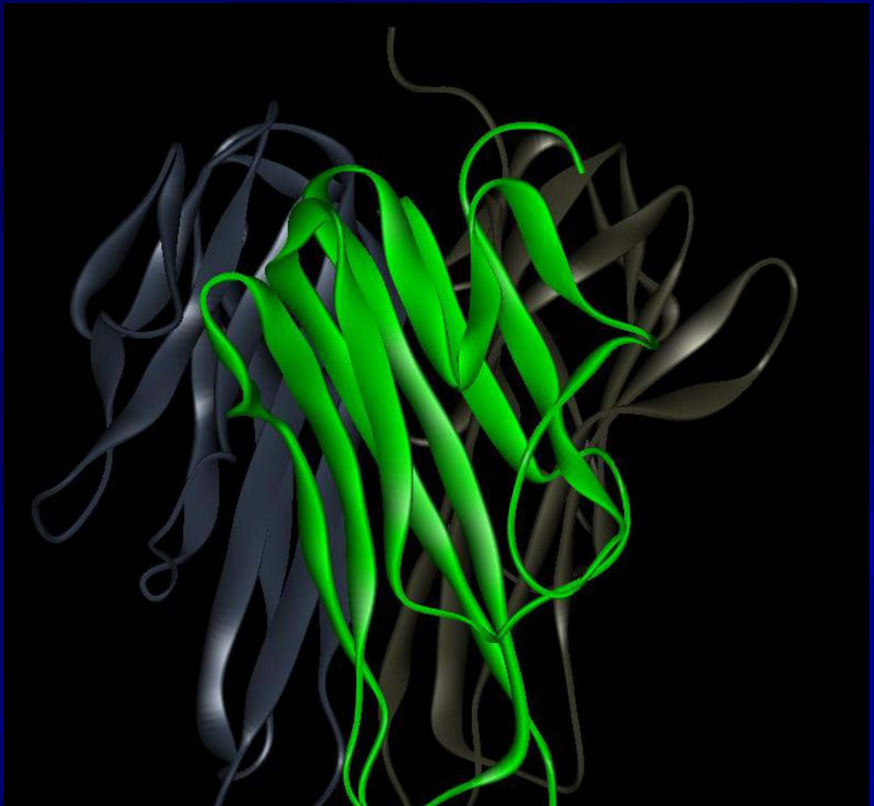
```
>PROTEIN  SEQUENCE
AQTVPYGIPL IKADKVQAQG FKGANVKVAV LDTGIQASHP DLNVVGGASF
VAGTNGMDVI NMSLGGASGS TAMKQAVDNA YARGVVVVAA AGNSGNSGST
NTIGYPAKYD SVIAVGAVDS NSNRASFSSV GAELEVMAPG AGVYSTYPTN
TYATLNGTSM ASPHVAGAAA LILSKHPNLS ASQVRNRLSS TATYLGSSFY
YGKGLINVEA AAQ*
```

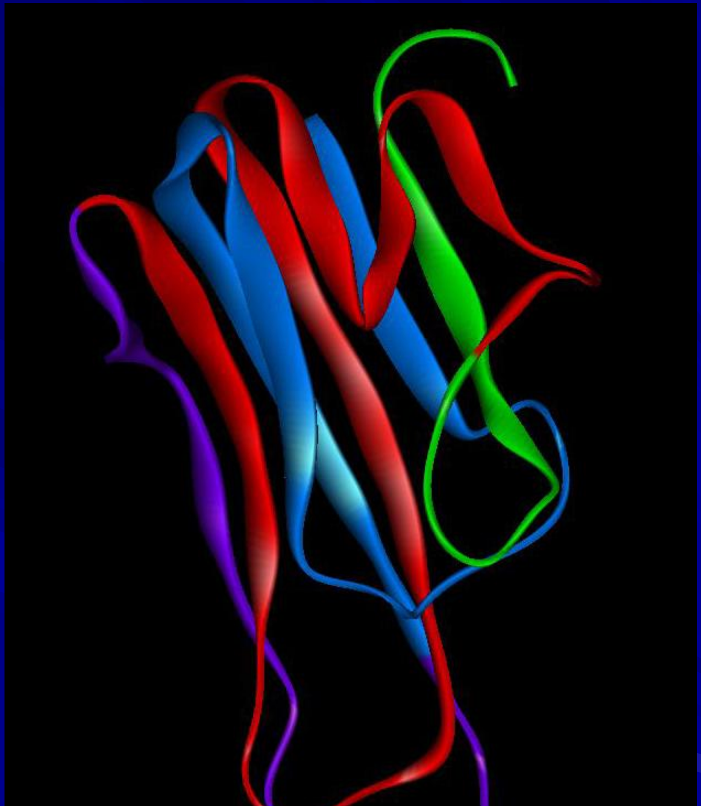
Biomol. Struct. Dyn. v.24(5), p.455

ПОЛУЧЕНИЕ НОВЫХ РЕКОМБИНАНТНЫХ ФОРМ IL-13

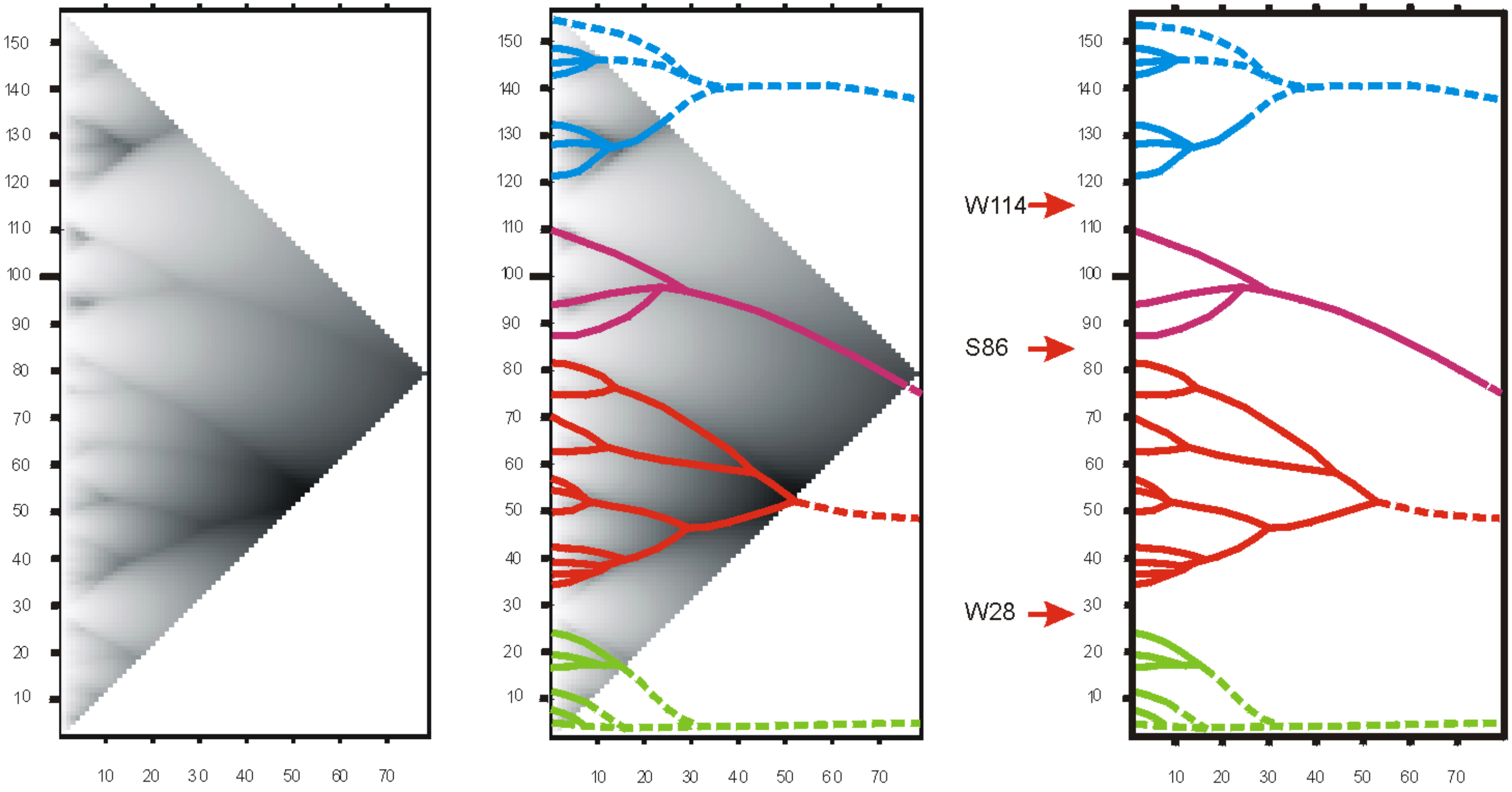


ПРОСТРАНСТВЕННАЯ СТРУКТУРА ФАКТОРА НЕКРОЗА ОПУХОЛЕЙ (TNF)





ИНФОРМАЦИОННАЯ СТРУКТУРА ФАКТОРА НЕКРОЗА ОПУХОЛЕЙ (TNF)



CONCLUSION

A NEW METHOD OF ANALYSIS OF PROTEIN SEQUENCES IS DEVELOPED. THE METHOD ALLOWS:

-TO DETECT THE HIERARCHICAL ORGANIZATION OF PROTEIN SEQUENCES.

- TO INVESTIGATE THE LATENT ELEMENTS OF THE SPATIAL ORGANIZATION OF PROTEINS WHICH CORRESPOND TO VARIOUS LEVELS OF HIERARCHY.

- TO DETECT NEW STRUCTURAL FACTORS INFLUENCE ON EFFICIENCY OF INTERACTIONS BETWEEN DIFFERENT POLYPEPTIDE CHAINS

E-mail:
alexei_nekrasov@mail.ru
alexei.n.nekrasov@gmail.com

FAX: +7 495 330 7103

PHONE: +7 926 119 34 91
(RUS)

THANK YOU !