NUMBERS

Estimated number of proteins in the human body: 100 000

Primary structure analysis (F. Sanger, 1953)

1953-1978 (25 years)	1081
1979-1991 (13 years)	16 000
1992-	1000/year

Three-dimensional (3D) structure (J. Kendrew, 1962)

1962-1985 (20 years)	200
1986-1991 (5 years)	480
1992-	100/years

CLASSIFICATION OF PROTEINS ACCORDING TO THEIR FUNCTION

1. Enzymatic catalysis (e.g. Ser proteases)
2. Transport (e.g. transferrin for iron, serum albumin for fatty acids)
3. Storage (e.g. ferrin for iron in liver, casein in milk)
4. Protection
    - toxins (e.g. ricin [plant], diphteria [bacteria])
    - self and non-self discrimination, immune protection (e.g. antibodies, antigenes)
5. Signal transduction (e.g. hormones, receptors)
    - nerve impulses
    - growth
    - differentiation
6. Cell to cell communication (e.g. adhesion, molecules; factors, acceptors)
7. Coordinated motion (e.g. muscle proteins)
8. Mechanical support
    - at cellular level (e.g. Membrane proteins)
    - at tissue level (e.g. structural proteins, e.g. collagen in skin, bone)

RECOGNITION PHENOMENA

Interaction	K_d[M]
1. Enzyme – substrate	10^{-3} – 10^{-5}
2. Transporter – ligand	10^{-6} – 10^{-8}
3. Hormone – receptor	10^{-9}
4. Antibody – antigen	10^{-7} – 10^{-11}
5. Storage protein – ligand	
6. Toxin – receptor	
7. Protein – protein (in a contractile super assembly)	
8. Lectin – carbohydrate	10^{-4} – 10^{-7}
9. Avidin – biotin	10^{-15}

METHODS FOR THE LOCALISATION OF FUNCTIONALLY RELEVANT DOMAINS IN PROTEINS

I. Experimental methods

Chemical modification
- side chain modification
- conjugation

Fragmentation
- enzymatic (e.g. trysin)
- chemical (e.g. BrCN)

Separation
- centrifugation
- chromatography
- electrophoresis

Identification
- amino acid analysis
- sequencing
- mass spectrometry

METHODS FOR THE LOCALISATION OF FUNCTIONALLY RELEVANT DOMAINS IN PROTEINS

I. Experimental methods

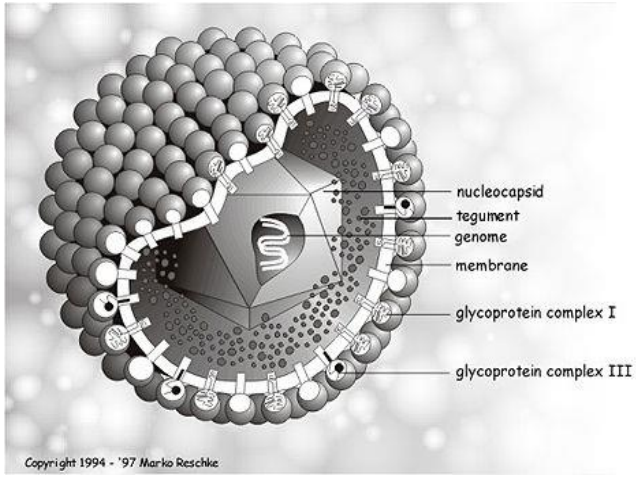
Chemical synthesis

- substituted analogs
- truncated/omitted analogs
- overlapping peptides
- peptide libraries

Genetic engineering

- deletion
- chimeric proteins
- mutagenesis
  - site directed
  - random
- phage display libraries

An example:
Identification of epitope sequences in protein antigens

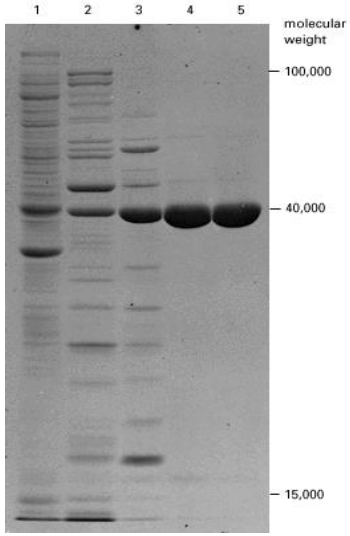


Whole virus

1. Affinity chromatography
2. Gradient centrifugation

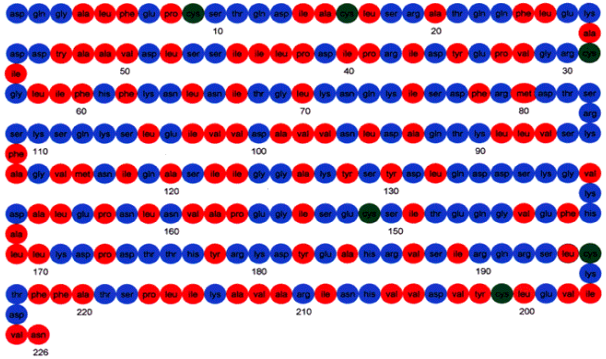
Viral envelope proteins
(mixture)

Immuneprecipitation
(Gelelectrophoresis)



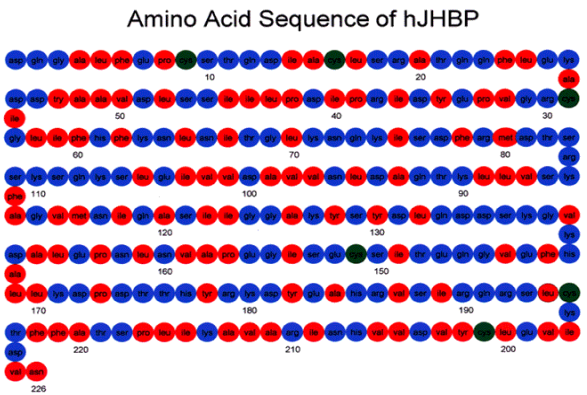
Immunodominant protein component komponens

Sequencing



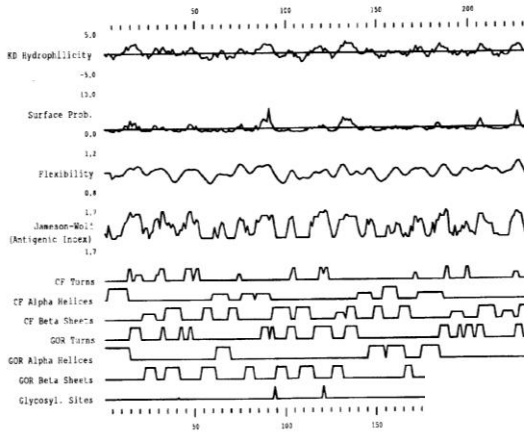
Amino acid sequence

Identification of peptid epitopes

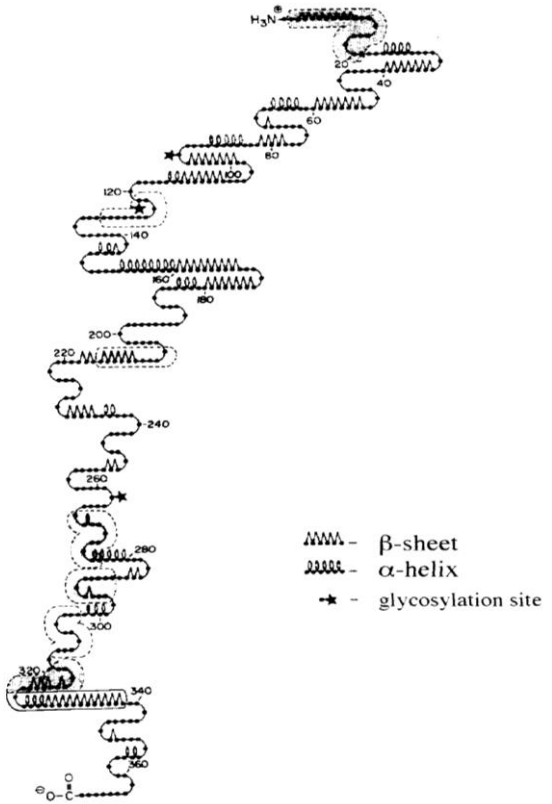


protein

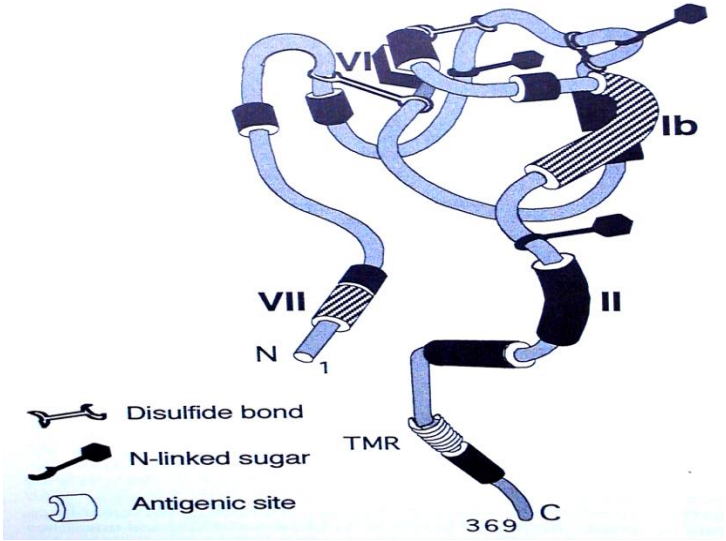
prediction of 3D structure



Prediction of hydrophilicity



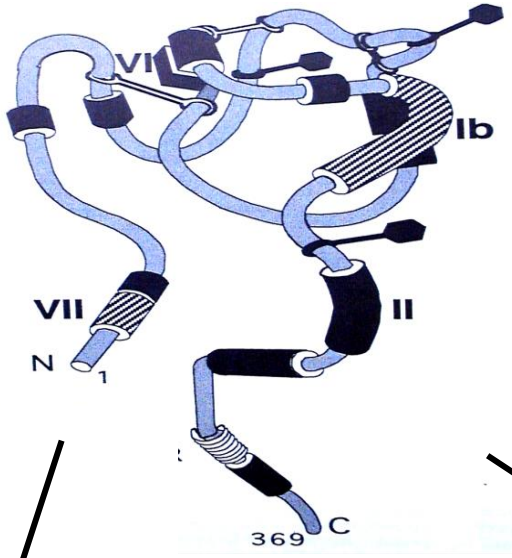
modeling

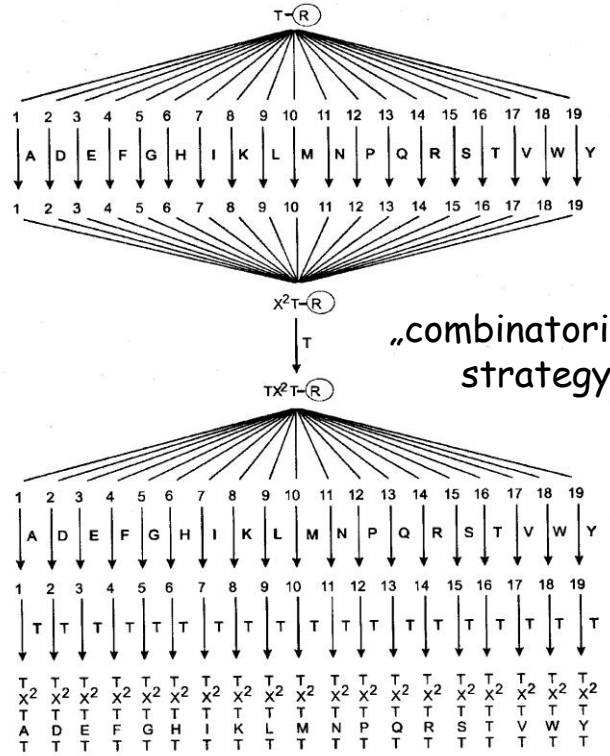


epitope „map"

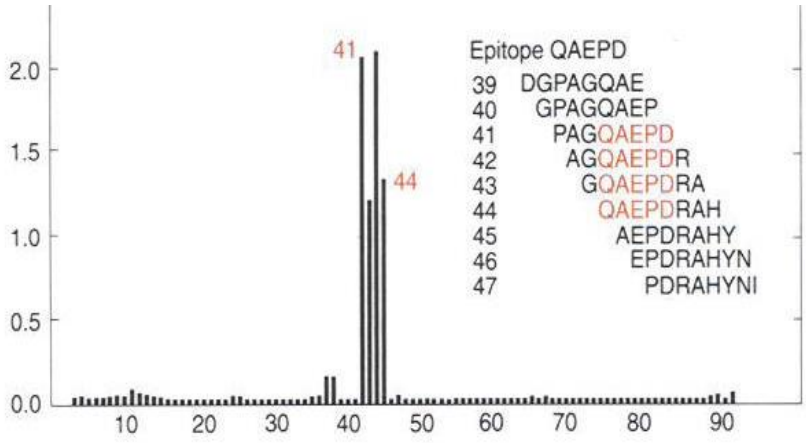
Identification of short sequences responsible for activity

„predicted"

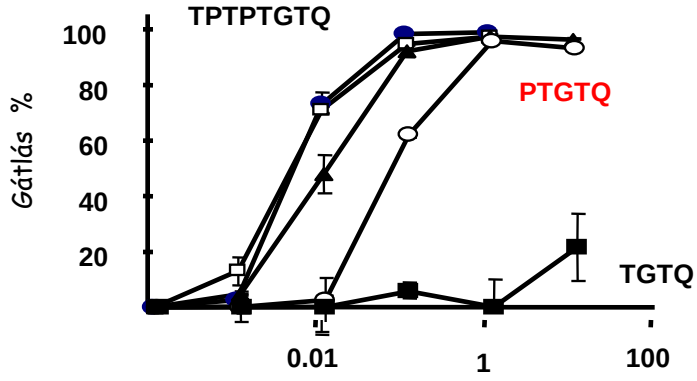




„combinatorial" strategy



„overlapping" strategy



„shortening" strategy

METHODS FOR THE LOCALISATION OF FUNCTIONALLY RELEVANT DOMAINS IN PROTEINS

II. Theoretical methods

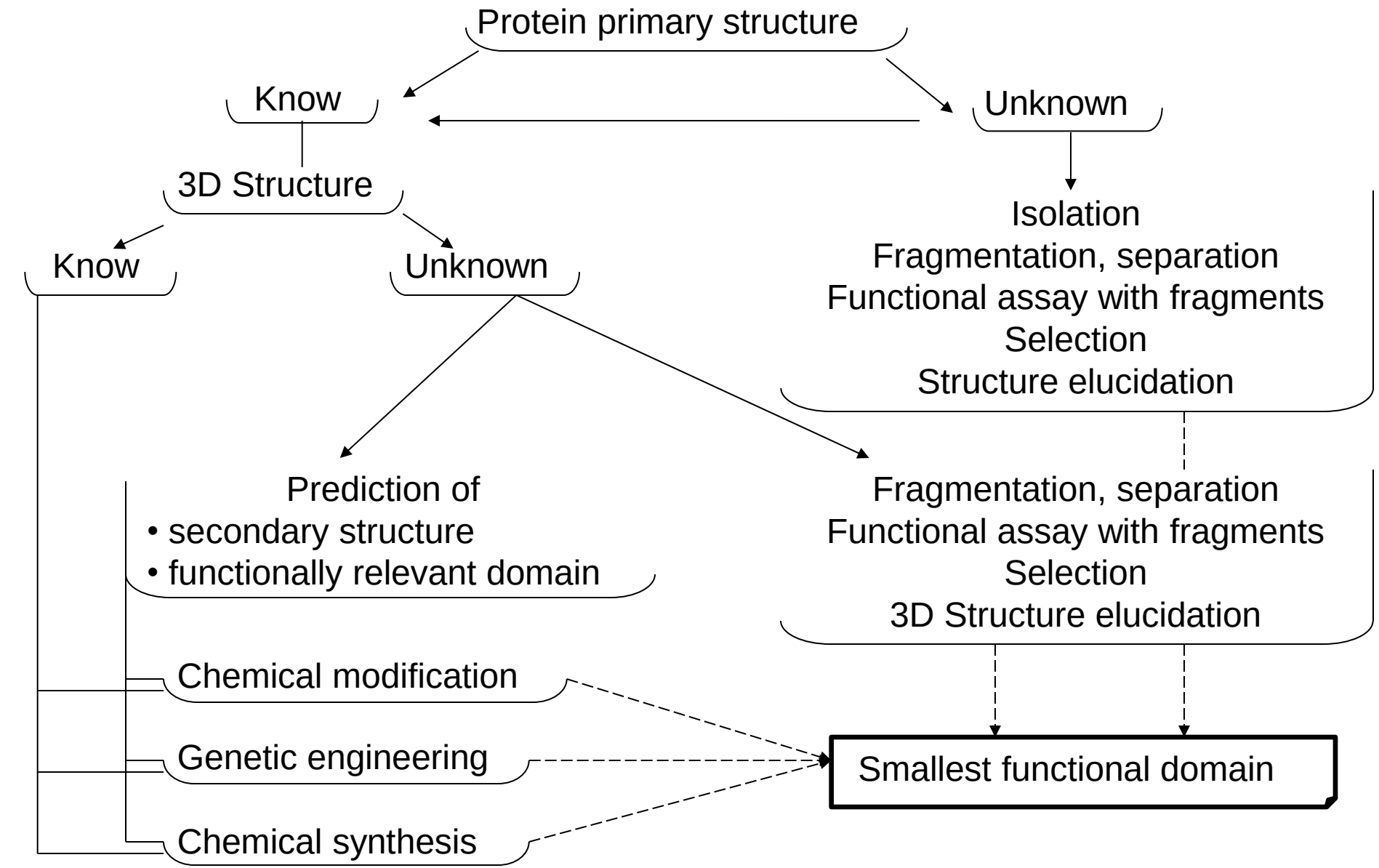
Quantum chemistry

- molecular mechanics
- molecular dynamics

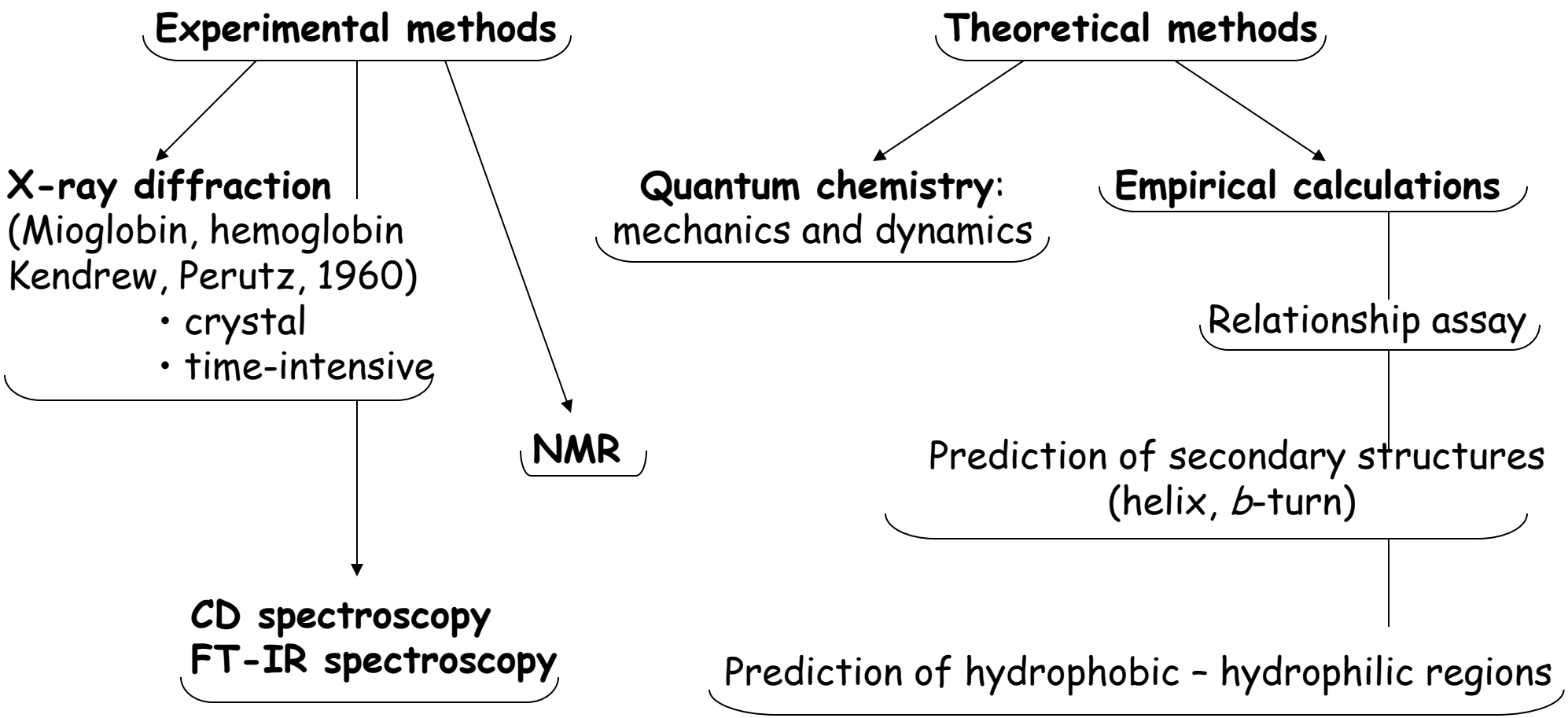
Predictions from the primary structure

- Probabilistic (statistical) 1970 -
- Physicochemical 1974 -
- Information theory 1974 -

Approaches for the localisation of functionally relevant domains in proteins



Strategies for determinations of 3D structures



Techniques for the detection of interaction/recognition phenomena

1. Molecular level

Detection with Separation

Separation techniques
- equilibrium dialysis
- chromatography
  - gel filtration
  - affinity
- electrophoresis

Detection techniqes
- spectroscopic
- radiochemical (^{125}I, ^{35}S, ^{3}H, ^{14}C)
- imunochemical
  - RIA/ELISA
  - blotting
  - immunprecipitation

Detection without Separation

Optical techniques
- absorption spectroscopy
- CD
- fluorescence spectroscopy
- IR and Raman spectroscopy

Resonance techniques
- NMR
- electron paramagnetic resonance (EPR)

Scattering and Diffraction techniques
- X-ray crystallography
- neutron scattering
- electron microscopy

Techniques for the detection of interaction/recognition phenomena

2. Cellular level

Bioassay (in vitro)
- binding to cell
- hemolysis
- antibacterial effect
- cytotoxicity

Example 1

Heparin binding domains of Apolipoprotein E

Heparin: glycosaminoglycan (*GAG*)

Apolipoprotein E: human plasma lipoprotein (ApoE)
299 amino acid, 3D structure unknown

Aim: Identification of primary *GAG* interaction sites (motifs),
which can be used for prediction based on initial sequence inspection

Phase I

1. step Fragmentation of the lipid-free protein

H----------Arg^{191}-Ala^{192}----------OH

+ thrombin

H ----------Arg^{191}-OH H-Ala^{192}----------OH

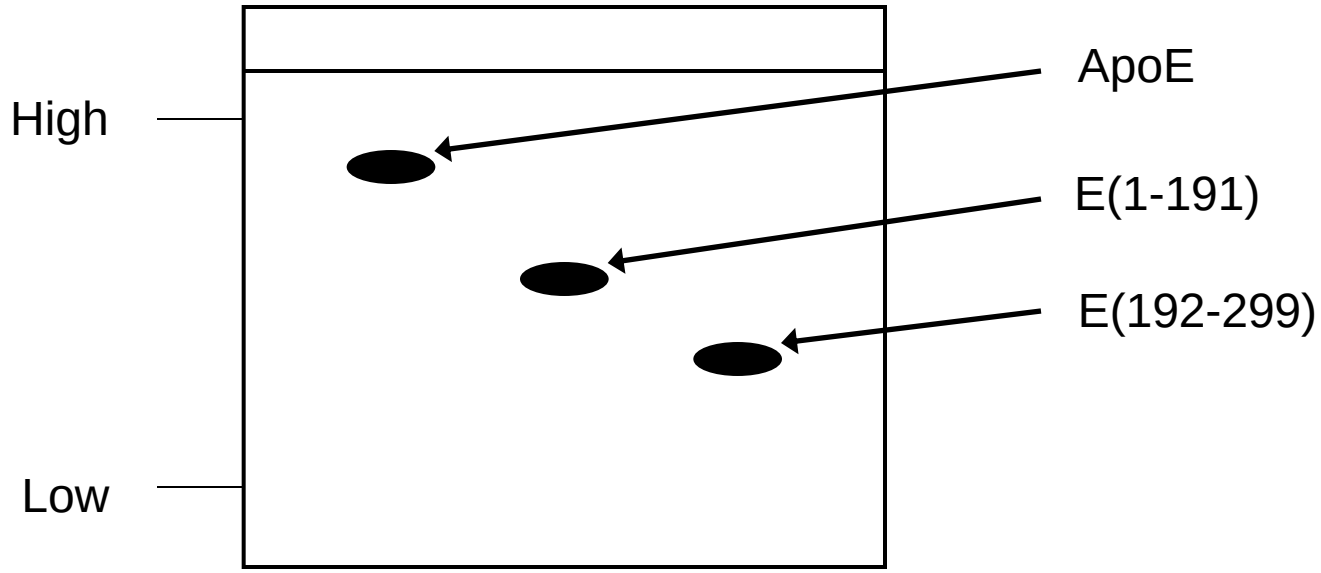
E(1-191) E(192-299)

Example 1

Heparin binding domains of Apolipoprotein E

Phase I

2. step Separation of the fragments



3. step Binding studies with ^{125}I-labelled heparin
   Method:
   a. Transfer of fragments to nitrocellulose by blotting
   b. Incubation of nitrocellulose with labelled heparin
   c. Radioautoradigraphy

Observation: ApoE and the two thrombin fragments bind heparin, indicating a minimum of two heparin-binding domains

Example 1

Heparin binding domains of Apolipoprotein E

Phase II

1. step Synthetic peptide with successive deletions from the amino- or/and carboxyl-terminal ends were prepared by SPPS

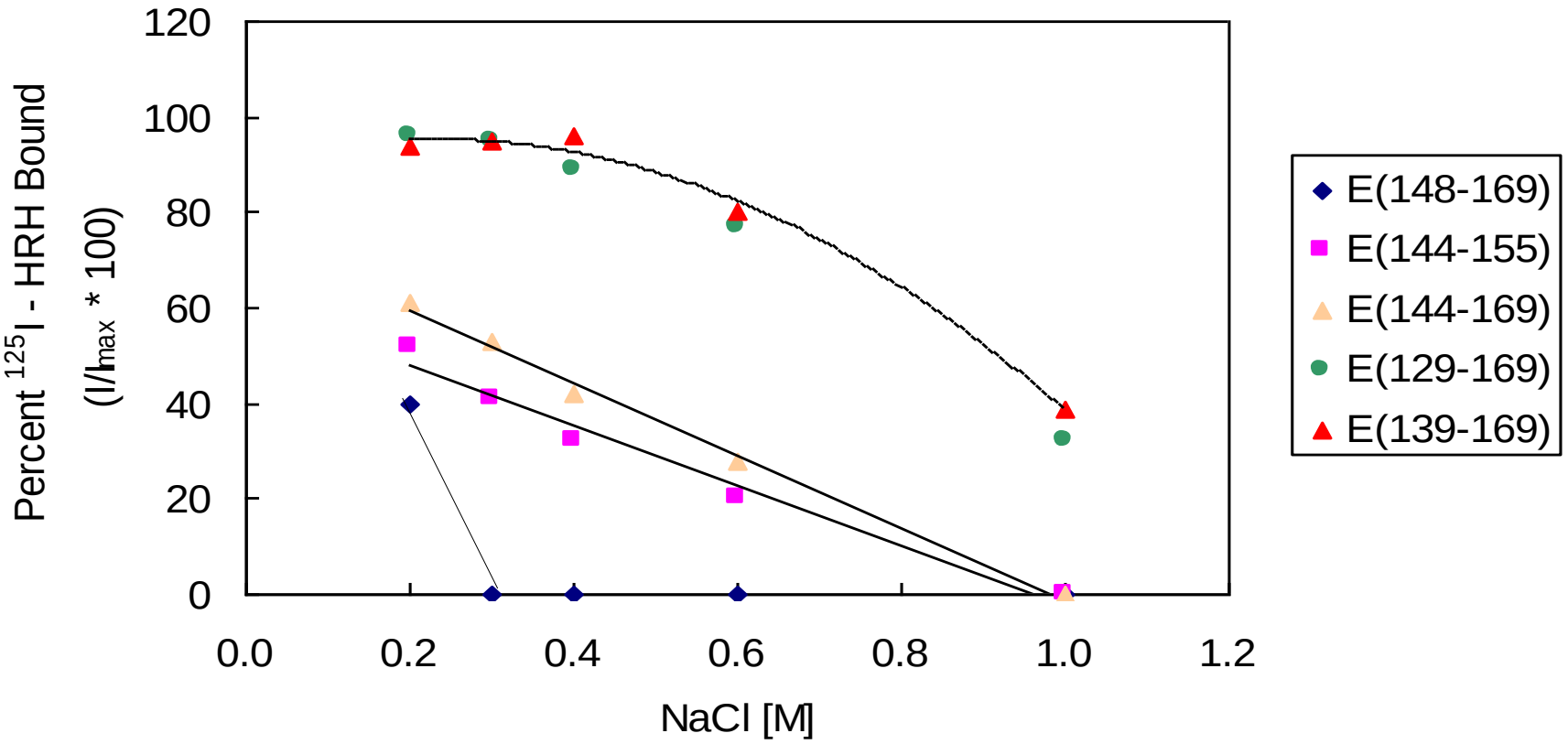
Residue	Amino acid sequence
129-169	STEELRVRLASHLRKLRKRLLRDADDLQKRLAVYQAGAREG
139-169	HLRKLRKRLLRDADDLQKRLAVYQAGAREG
144-169	LRKRLLRDADDLQKRLAVYQAGAREG
148-169	LLRDADDLQKRLAVYQAGAREG
141-155	LRKRLLRDADDL

Example 1

Heparin binding domains of Apolipoprotein E

Phase II

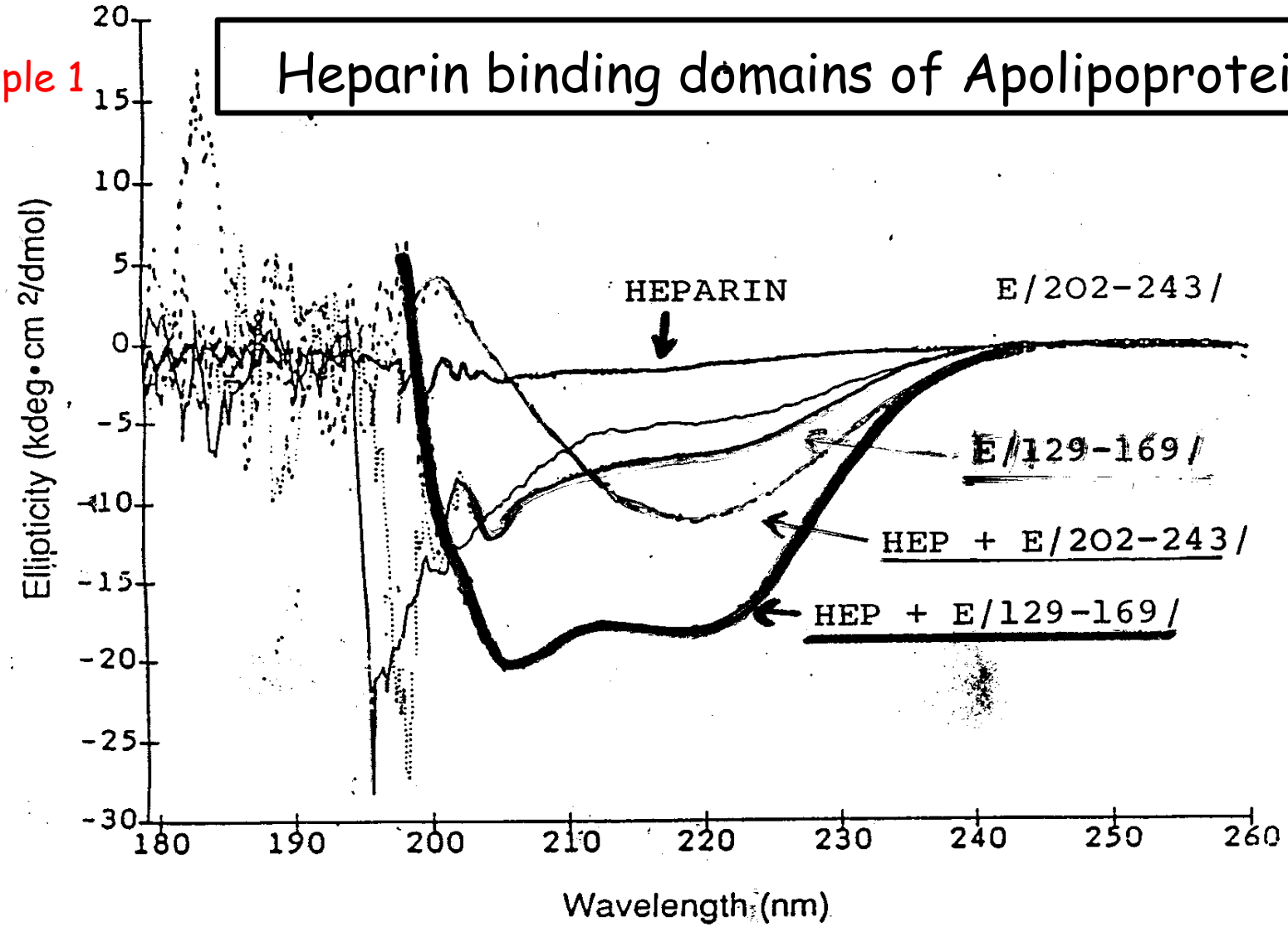
2. step Binding studies with ^{125}I-labelled heparin. Dot-blot assay.



Observation: The critical residues for heparin binding resided between amino acid 144 and 150 (Leu-Arg-Lys-Arg-Leu-Leu-Arg)

Example 1

Heparin binding domains of Apolipoprotein E



Effect of heparin on the circular dichroism of E(129–169) and E(202–243). The spectra represent the following: (a) 100 μg/ml E(129–169); (b) 100 μg/ml E(129–169) + 60 μg/ml heparin; (c) 100 μg/ml E(202–243); (d) 100 μg/ml E(202–243) + 60 μg/ml heparin; and (e) 60 μg/ml heparin.

CIRCULAR DICHROISM (CD) PARAMETERS AND SECONDARY STRUCTURES FOR APOE HEPARIN-BINDING PEPTIDES

Peptides	$[\Theta]_{208}$[a]	$[\Theta]_{222}$[a]	Determined by CD[b]		Predicted[c]		Simulated[d]	
			α Helix (%)	β Strand (%)	α Helix (%)	β Strand (%)	α Helix (%)	β Strand (%)
ApoE(129–169)								
− Heparin	−4,590	−2,448	22	5	—	—	25 (40)	
+ Heparin	−19,500	−17,500	49	<1	63	17	56 (65)	—
ApoE(202–243)								
− Heparin		−4,200	11	16	—	—	—	13 (26)
+ Heparin		−12,000	9	67	40–60	—	—	32 (48)

[a] deg cm²/dmol.

[b] Determined by deconvolution of CD spectra.[20]

[c] Values were determined by the predictive method of Chou and Fasman.[11]

[d] Simulations were determined by molecular dynamics calculations as described in the text for peptides E(129–159) (α helix) and E(211–234) (β strand) complexed to a heparin octasaccharide and hexadecasaccharide, respectively. For the E(211–234) and E(129–159) studies, two (in parentheses) and three residue windows were used to define β-strand and α-helical character, respectively. The values represent the results averaged over the last 20 psec of the 80-psec simulations (60 psec $\leq \tau \leq$ 80 psec). See text for details.

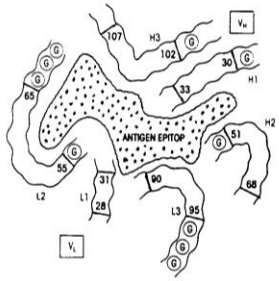
Example 2

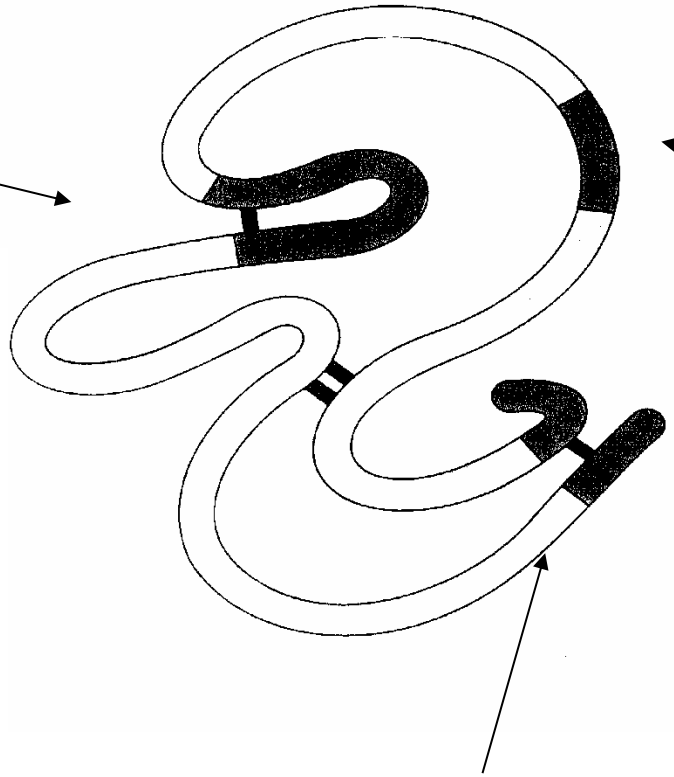
Molecular Mapping of Immunogenic Determinants of Human CD4 Using Chimeric Interspecies Molecules and Antibodies

CD4/L3T4:	lymphocyte antigen, expressed on helper T-cells and macrophages, primary and 3D structure known for CD4 (human), primary structure known for L3T4 (mouse)
Antibodies:	37 human and mouse monoclonal antibodies recognizing CD4/L3T4 positive cells
Interaction:	Antibody – antigen interactions
Aim:	Identification of epitope regions recognized by anti-CD4 antibodies. (Lack of binding of certain antibodies to overlapping peptides corresponding to CD4 indicated the presence of discontinuous conformational epitopes.)

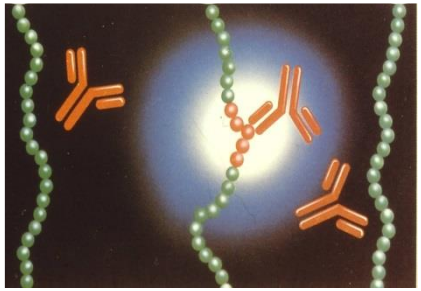
Peptide epitopes – epitope recognition

Topographic, continous
(antibody epitope)

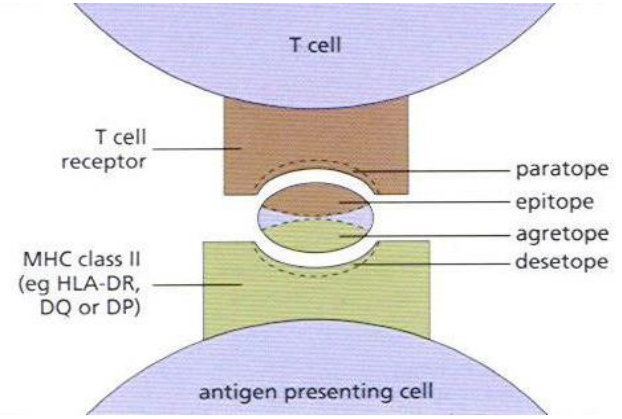




Linear, sequential
(antibody or T-cell epitope)



Topographic,non -continuous
(antibody epitope)



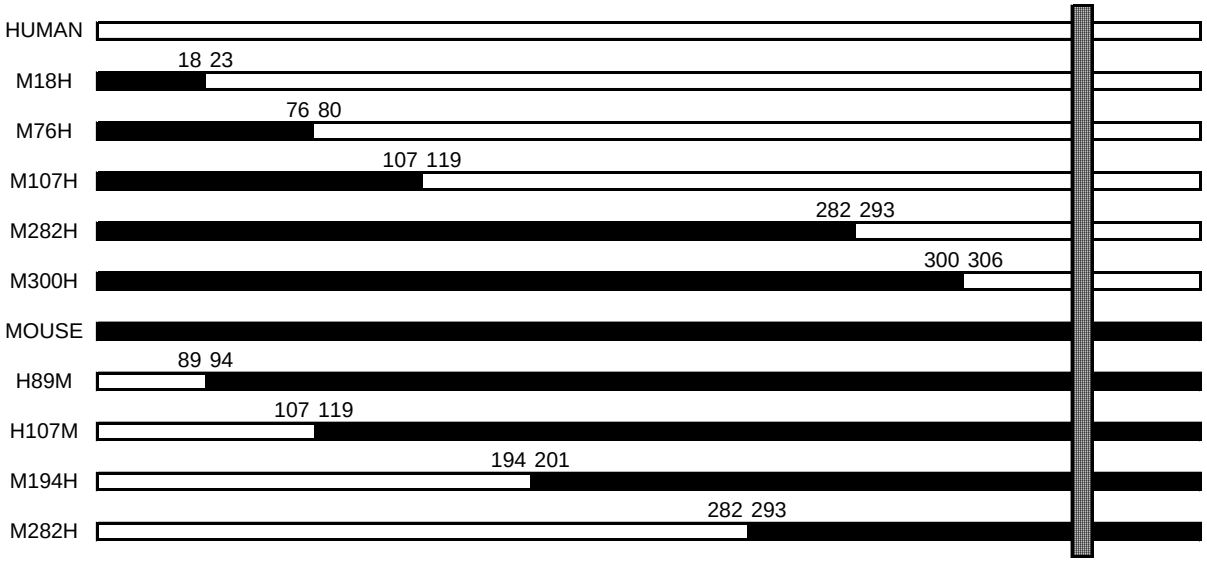
Example 2

Molecular Mapping of Immunogenic Determinants of Human CD4 Using Chimeric Interspecies Molecules and Antibodies

1. step Preparation of chimeric CD4 cDNA molecules using human CD4 and mouse L3T4 cDNA clones.
   Method: Recombinant DNA technique (Bacterial homologous recombination system)

2. step Expression of chimeric CD4 cDNA molecules
   Method: Transfection into an L3T4 negative variant of the murine T-cell line, EL-4

Observation: 9 chimeric L3T4/CD4 (mouse amino terminal) and CD4/L3T4 (human amino terminal) molecules with cross-over in the extracellular region of the mature protein were generated



	GROUP							
	A	B	C	D	E	F	rgp120	GK1.5
HUMAN	+	+	+	+	+	+	+	-
M18H	-	+	+	+	+	+	+	-
M76H	-	-	-	+	+	+	-	-
M107H	-	-	-	-	+	+	-	+
M282H	-	-	-	-	+	-	-	+
M300H	-	-	-	-	+	-	-	+
MOUSE	-	-	-	-	-	-	-	+
H89M	-	+	-	+	-	-	+	-
H107M	-	+	-	+	-	-	+	-
M194H	+	+	+	+	-	-	+	-
M282H	+	+	+	+	-	-	+	-

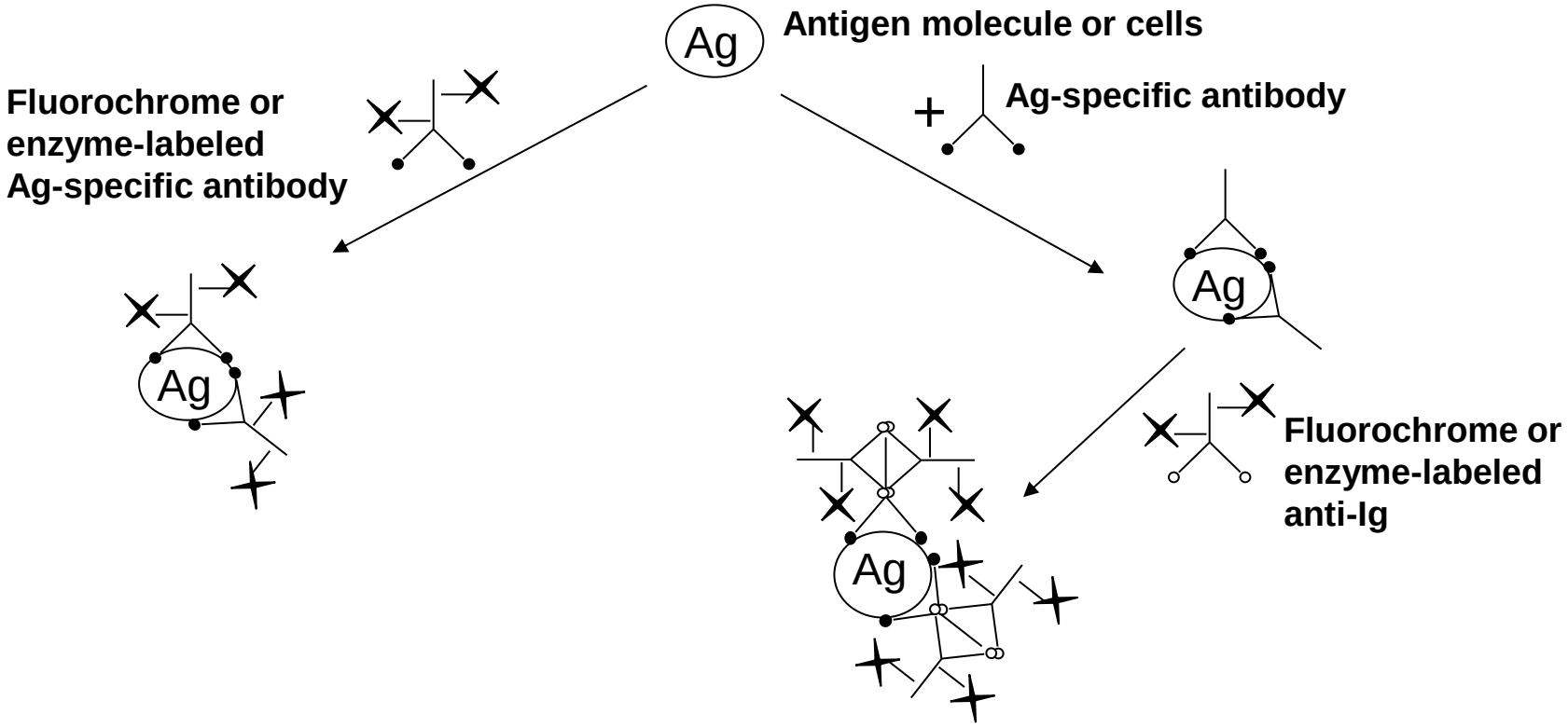
Example 2

Molecular Mapping of Immunogenic Determinants of Human CD4 Using Chimeric Interspecies Molecules and Antibodies

3. step Binding studies with purified CD4 or L3T4 specific antibodies on EL-4 cells expressing chimeric CD4 in 3D form.

   Method:

   a. Staining cells with CD4/L3T4 antibodies
   b. Incubation of stained cells with FITC-labeled goat anti-mouse IgG
   c. Flow cytometry on FACS



Example 2

Molecular Mapping of Immunogenic Determinants of Human CD4 Using Chimeric Interspecies Molecules and Antibodies

Observation: All chimeric molecules analysed in transfectants detectable with human and/or mouse specific anti-CD4 antibodies. Using the chimerics, it was possible to localize most of the CD4 epitopes to specific region of the CD4 protein.

NB.:
1. CD4/L3T4 recognize antigen in the context of class II MHC antigens.
2. As expected from their functional similarities, the human and the mouse CD4 molecules are highly homologous at both the DNA (70%) and amino acid (54%) levels.

Reference: P. Estess et al. Current Research in Protein Chemistry (ED.: J. J. Villafranca, Academic Press, San Diego, p. 499 (1990))

Example 3

Localization of Immunogenic Determinants (Epitopes) of human epithelial mucin glycoprotein, MUC-1 Using synthetic Peptides and MUC-1 specific Antibodies

MUC-1: high molecular mass, MUC-1 gene related glycoprotein, associated with human breast and ovarian carcinoma, primary structure is known

Antibodies: mouse monoclonal antibodies recognizing MUC-1 glycoprotein [HMFG-1, C595, B55, etc.]

Interaction: Antibody – antigen interactions

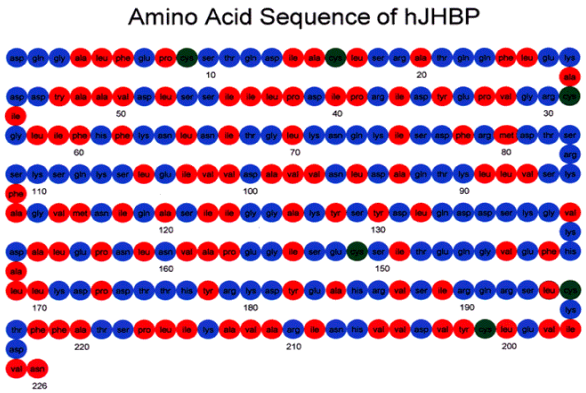
Aim: Identification of epitopes recognized by anti-MUC-1 antibodies

Phase I

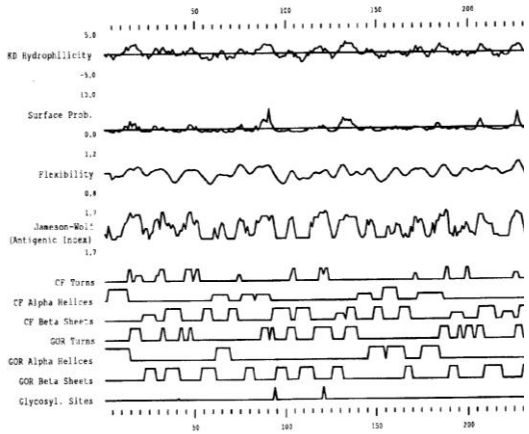
1. step Analysis of the primary structure of MUC-1 glycoprotein.

Method: Prediction of B-cell epitopes using various algorithms searching for
a. hydrophilic region and
b. β-turn secondary structure

Analysis of the primary structure

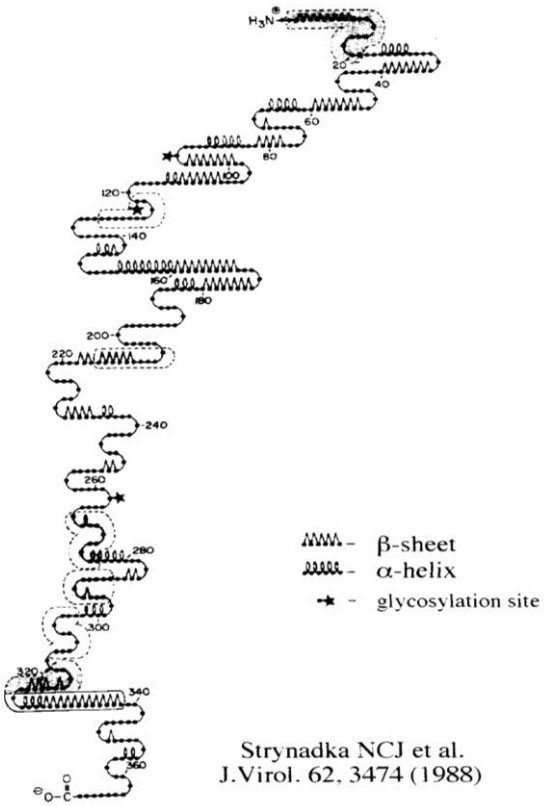


protein



prediction of 3D structure

Prediction of hydrophilicity



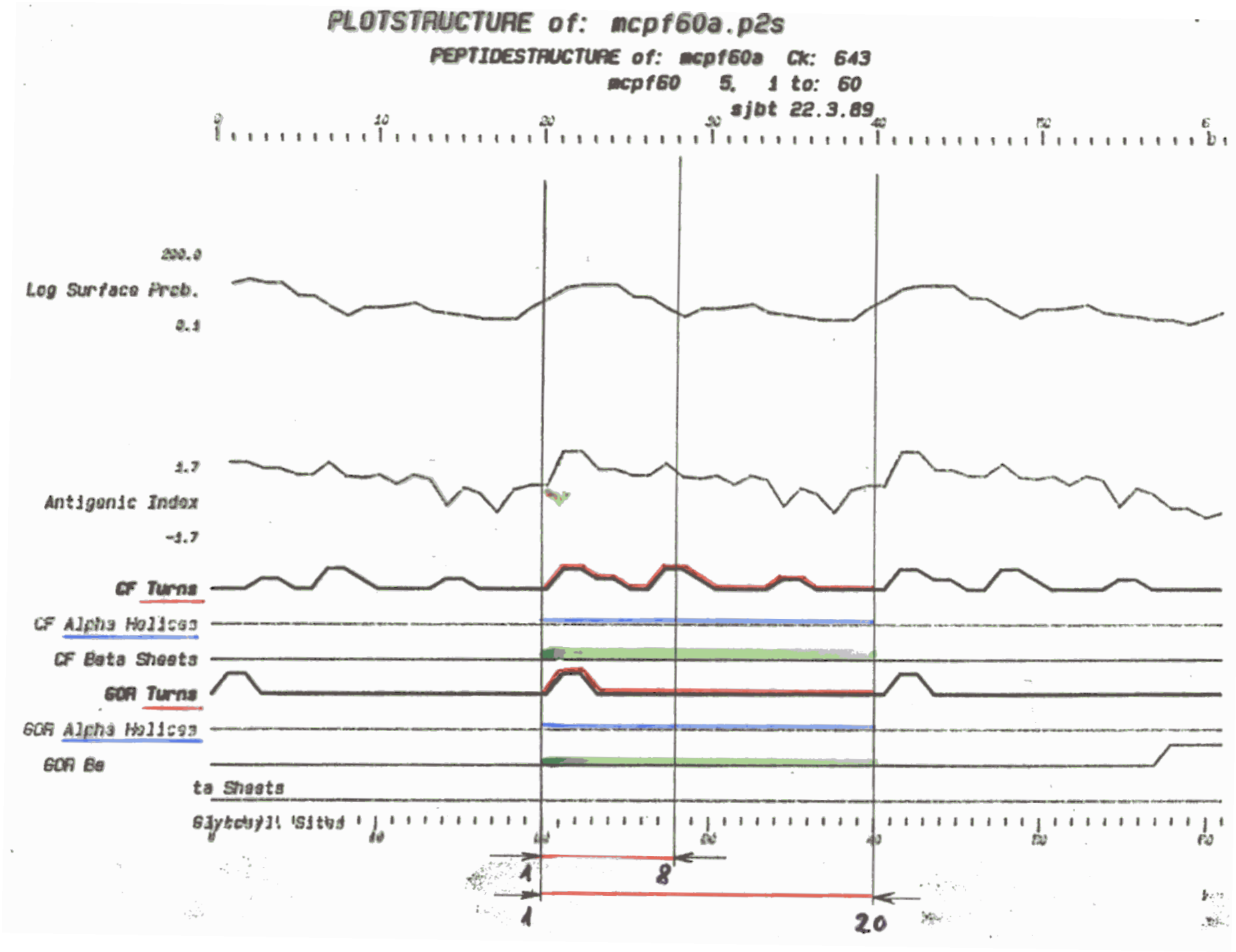
Hydrophobicity scales for amino acid residues[1]

Amino acid	Consensus (67)	Von Heijne (66)	Janin (62)	Chothia (63)	Wolfenden (61)	Tanford/Segrest (64, 65)	Kyte (68)	Argos (69)
Ile	0.73	4.4	0.7	0.24	2.15	5.0	4.5	1.67
Phe	0.61	5.2	0.5	0.0	-0.76	5.0	2.8	2.03
Val	0.54	3.9	0.6	0.09	1.99	3.0	4.2	1.14
Leu	0.53	4.2	0.5	-0.12	2.28	3.5	3.8	2.93
Trp	0.37	3.9	0.3	-0.59	-5.88	6.5	-0.9	1.08
Met	0.26	2.1	0.4	-0.24	-1.48	2.5	1.9	2.96
Ala	0.25	2.9	0.3	-0.29	1.94	1.0	1.8	1.56
Gly	0.16	1.9	0.3	-0.34	2.39	0.0	-0.4	0.62
Cys	0.04	-0.08	0.9	0.0	-1.24	0.0	2.5	1.23
Tyr	0.02	3.6	-0.4	-1.02	-6.11	4.5	-1.3	0.68
Pro	-0.07	1.1	-0.3	-0.90	---	1.5	-1.6	0.76
Thr	-0.18	1.2	-0.2	-0.71	-4.88	0.5	-0.7	0.91
Ser	-0.26	0.36	-0.1	-0.75	-5.06	-0.5	-0.8	0.81
His	-0.40	-1.5	-0.1	-0.94	-10.27	1.0	-3.2	0.29
Glu	-0.62	-4.0	-0.7	-0.90	-10.20	---	-3.5	0.23
Asn	-0.64	-1.0	-0.5	-1.18	-9.68	-1.5	-3.5	0.27
Gln	-0.69	-0.52	-0.7	-1.53	-9.38	-1.0	-3.5	0.51
Asp	-0.72	-5.6	-0.6	-1.02	-10.95	---	-3.5	0.14
Lys	-1.1	-2..3	-1.8	-2.05	-9.52	---	-3.9	0.15
Arg	-1.8	-9.4	-1.4	-2.71	-19.92	---	-4.5	0.45

[1]The order is by decreasing hydrophobicity on the consensus scale. The magnitudes for all but the two scales on the right may be considered Roughly in kcal*mol^{-1} transfer from a hydrophobic to a hydrophilic phase. The scales do not all measure the same property

Hydrophobicity scales for amino acid residues

Amino acid	Kyte and Doolittle	Eisenberg et al.
Isoleucine	4.5	0.73
Valine	4.2	0.54
Leucine	3.8	0.53
Phenylalanine	2.8	0.61
Cysteine / cistin	2.5	0.04
Methionine	1.9	0.26
Alanine	1.8	0.25
Glycine	-0.4	0.16
Threonine	-0.7	-0.18
Tryptophan	-0.9	0.37
Serine	-0.8	-0.26
Tyrozine	-1.3	0.02
Proline	-1.6	-0.07
Histidine	-3.2	-0.40
Glutamic acid	-3.5	-0.62
Glutamine	-3.5	-0.69
Aspartic acid	-3.5	-0.72

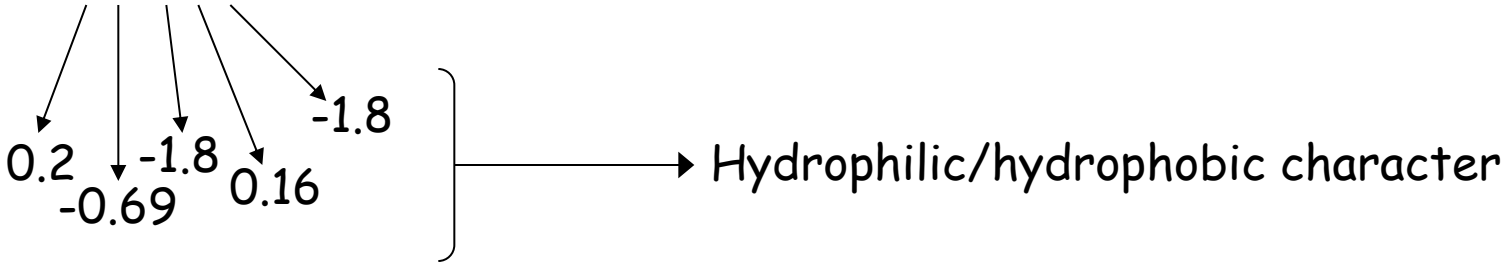

PLOTSTRUCTURE of: mcpf60a.p2s
PEPTIDESTRUCTURE of: mcpf60a Ck: 643
mcpf60 5. 1 to: 60
sjbt 22.3.89
200.0
Log Surface Prob.
0.1
1.7
Antigenic Index
-1.7
CF Turns
CF Alpha Helices
CF Beta Sheets
GOR Turns
GOR Alpha Helices
GOR Be
ta Sheets
Glycosyl Sites
1
8
1
20

Example 4

Location of membrane-spanning region(s) of human intracellular adhesion molecule 1 (ICAM)

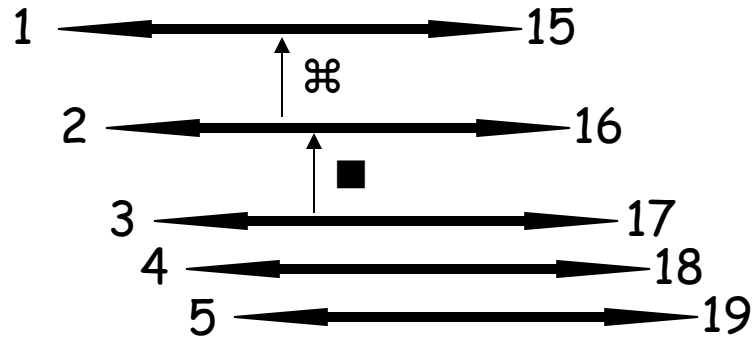
1. step Assignment

^{I}TQRGRFDAPTPTPTGTQTPTTDASERTHILEPEAKNADGTREQWFY^{46}....



2. step Averaging

^{I}TQRGRFDAPTPTPTG^{15}TQTPTTDASERTHILEPEAKNADGTREQWFY^{46}....



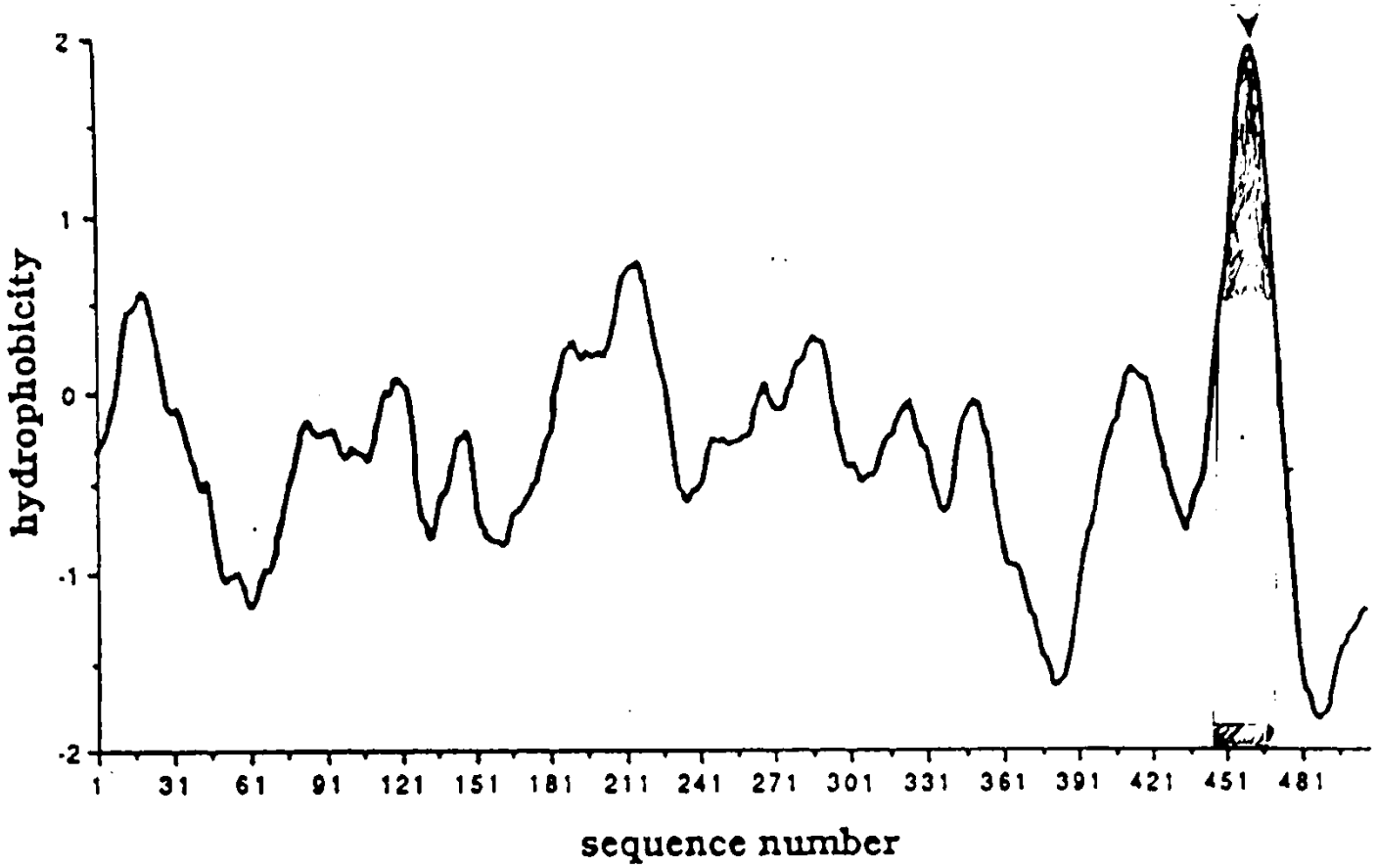
$⌘ = (n_1+n_2+n_3+ \ldots\ldots +n_{15}) : 15$

$■ = (n_2+n_3+n_4+ \ldots\ldots +n_{16}) : 15$

Example 4

Location of membrane-spanning region(s) of human intracellular adhesion molecule 1 (ICAM)

3. step Evaluation



Simulation of receptor environment