united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR 1329 - 1

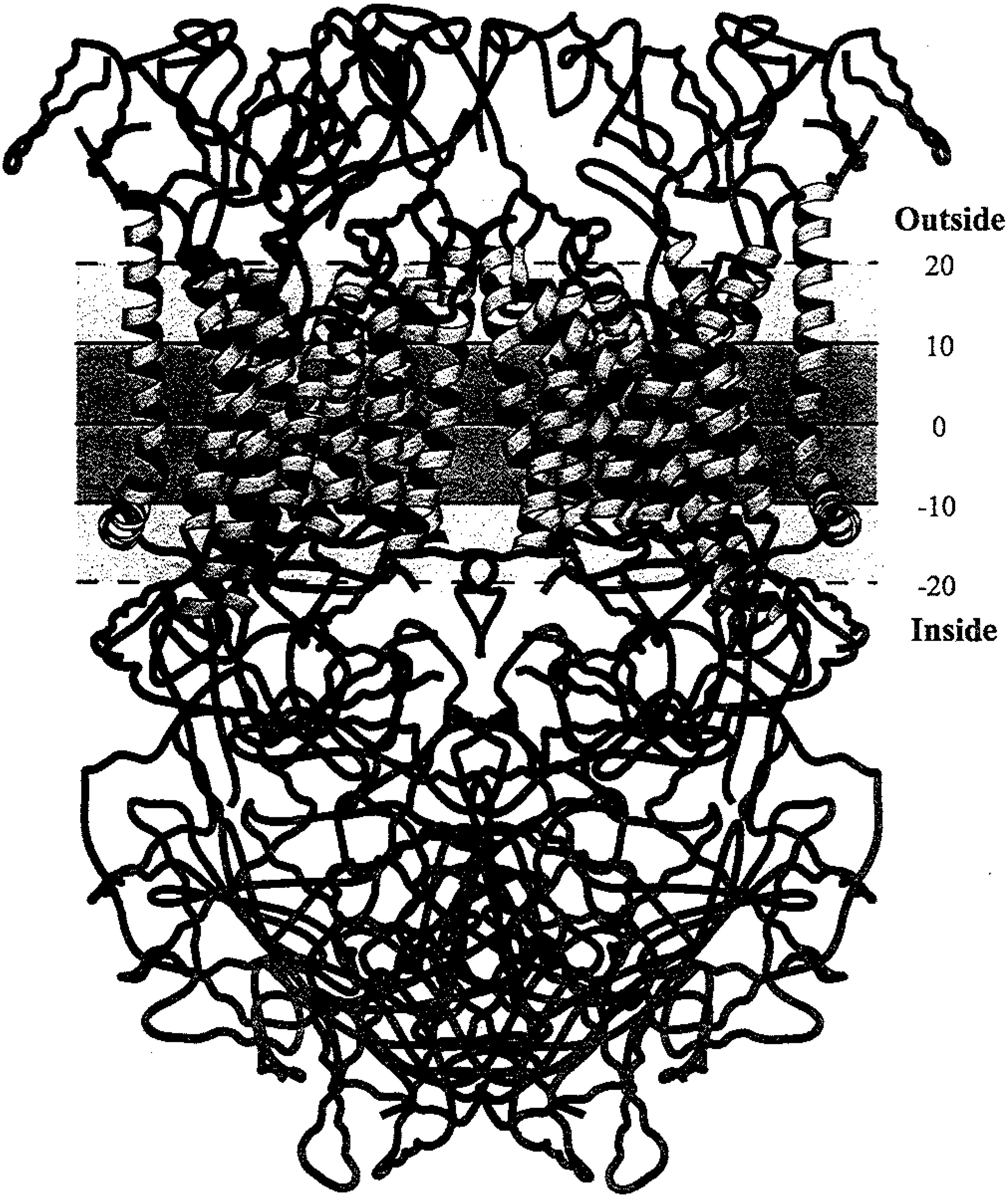
**COLLEGE ON BIOPHYSICS:
FROM MOLECULAR GENETICS TO STRUCTURAL BIOLOGY**

1 - 12 OCTOBER 2001

PREDICTION OF TRANSMEMBRANE PREOTEIN TOPOLOGY

Istvan SIMON
Institute of Enzymology, BRC
Hungarian Academy of Sciences
XI. Karolina uut. 29
H-1518 Bp, Budapest, Hungary

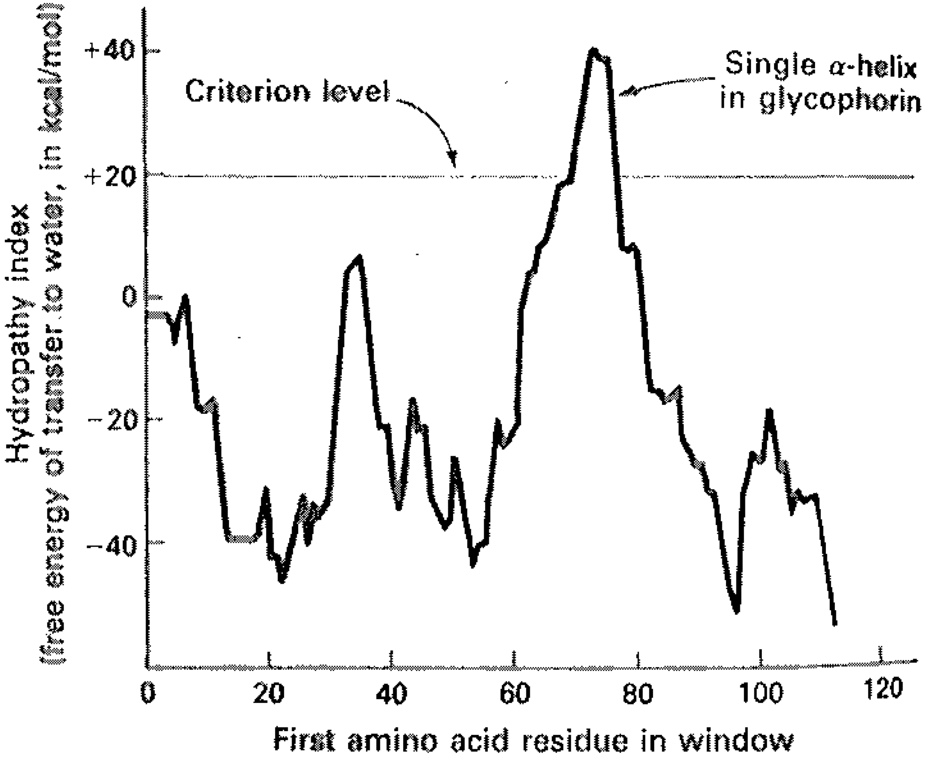
Cytochrome BC1 Complex (1bgy)

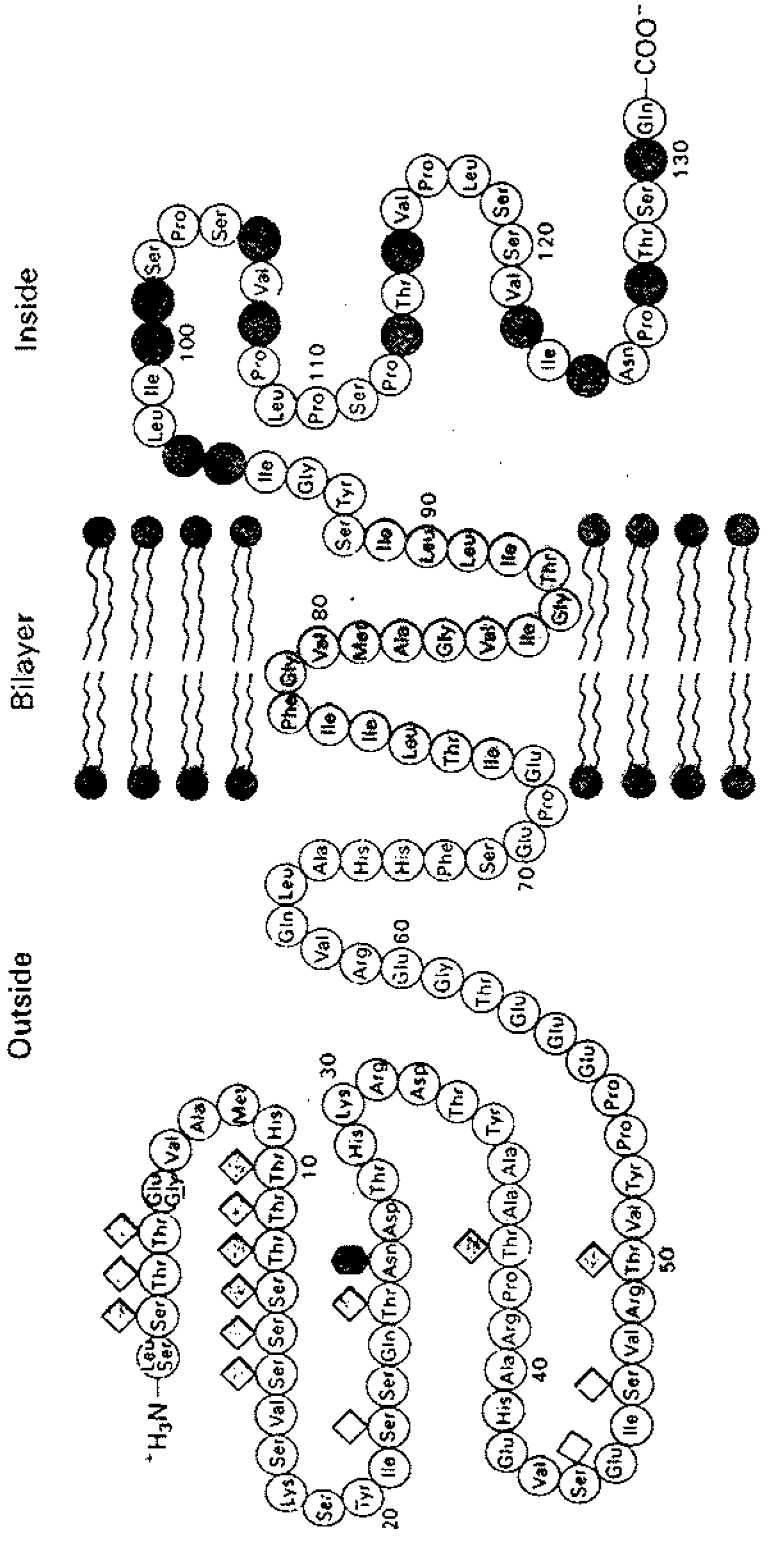


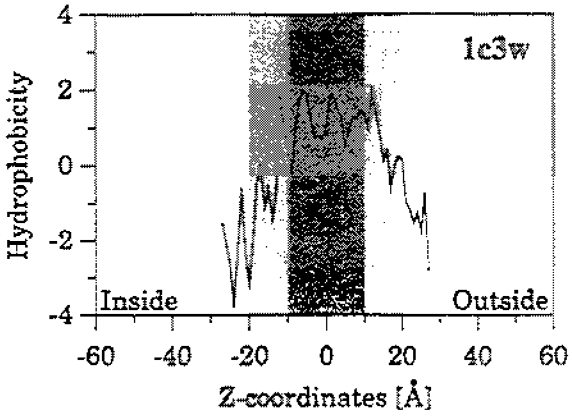
Polarity scale for identifying transmembrane helices

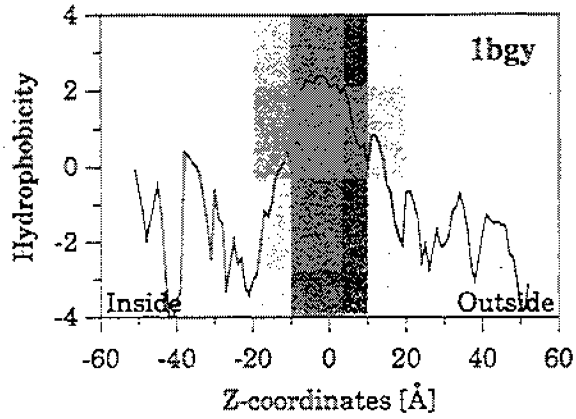
Amino acid residue	*Transfer free energy (kcal/mol)*
Phe	3.7
Met	3.4
Ile	3.1
Leu	2.8
Val	2.6
Cys	2.0
Trp	1.9
Ala	1.6
Thr	1.2
Gly	1.0
Ser	0.6
Pro	−0.2
Tyr	−0.7
His	−3.0
Gln	−4.1
Asn	−4.8
Glu	−8.2
Lys	−8.8
Asp	−9.2
Arg	−12.3

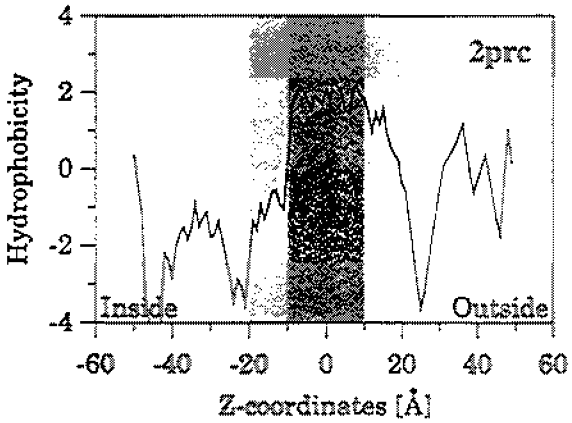
Note: The free energies are for the transfer of an amino acid residue in an α helix from the membrane interior (assumed to have a dielectric constant of 2) to water. After D.M. Engelman, T.A. Steitz, and A. Goldman. *Ann. Rev. Biophys. Biophys. Chem.* 15(1986):330.

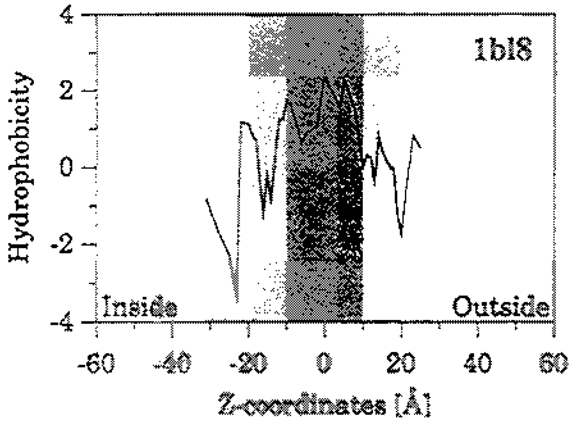


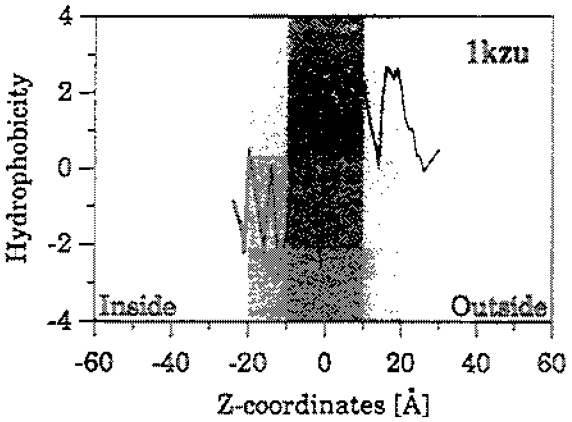

Outside
Bilayer
Inside
^+H_3N
COO^-
10
20
30
40
50
60
70
80
90
100
110
120
130

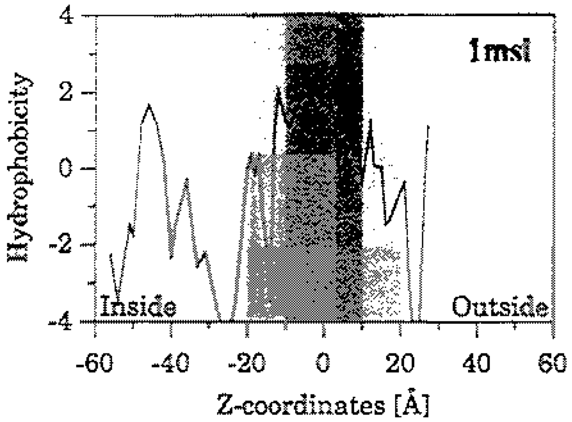

1c3w
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]

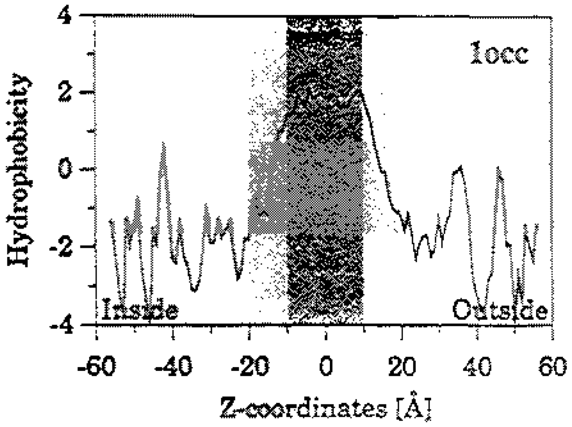

1bgy
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]

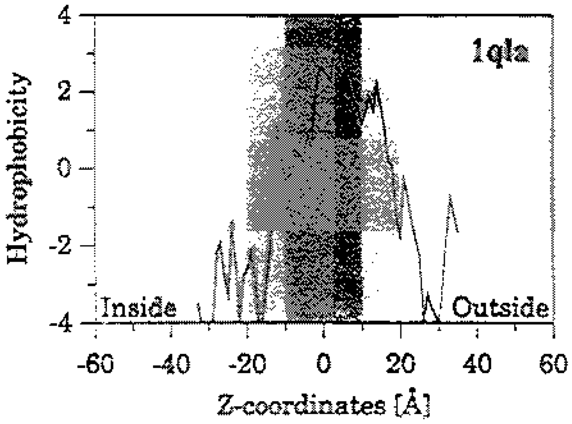

2prc
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]


1bl8
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]


1kzu
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]


1msl
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]


1occ
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]


1qla
Hydrophobicity
4
2
0
-2
-4
Inside
Outside
-60
-40
-20
0
20
40
60
Z-coordinates [Å]

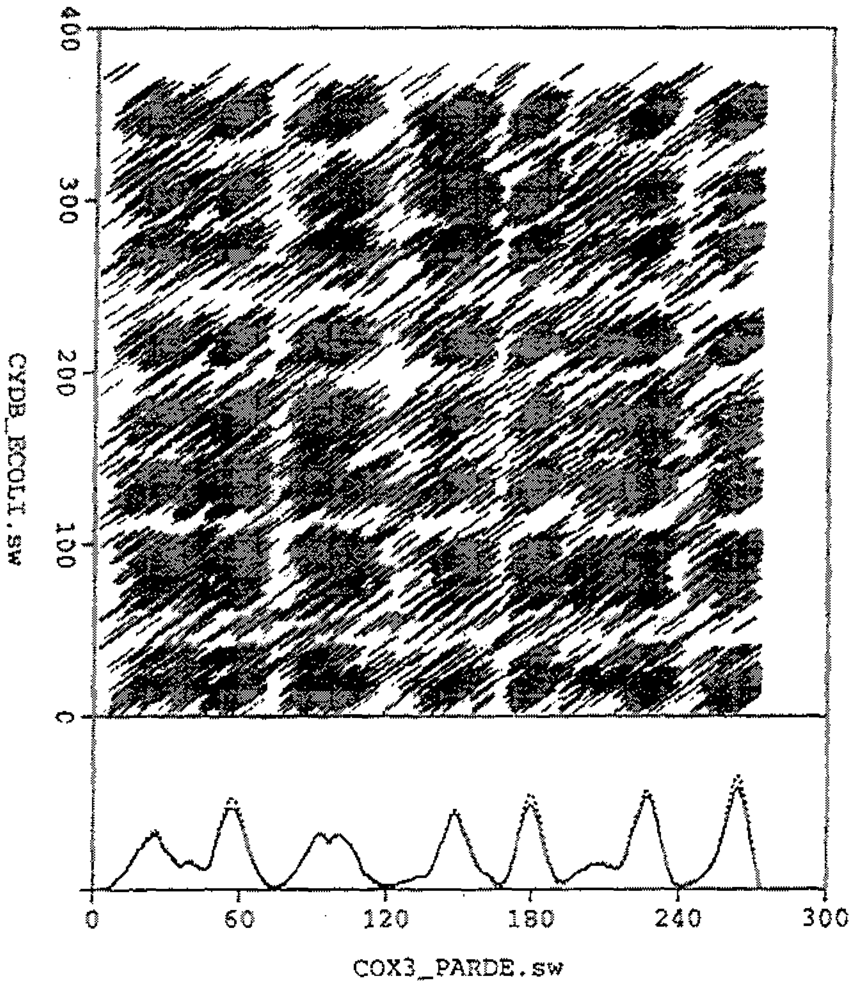
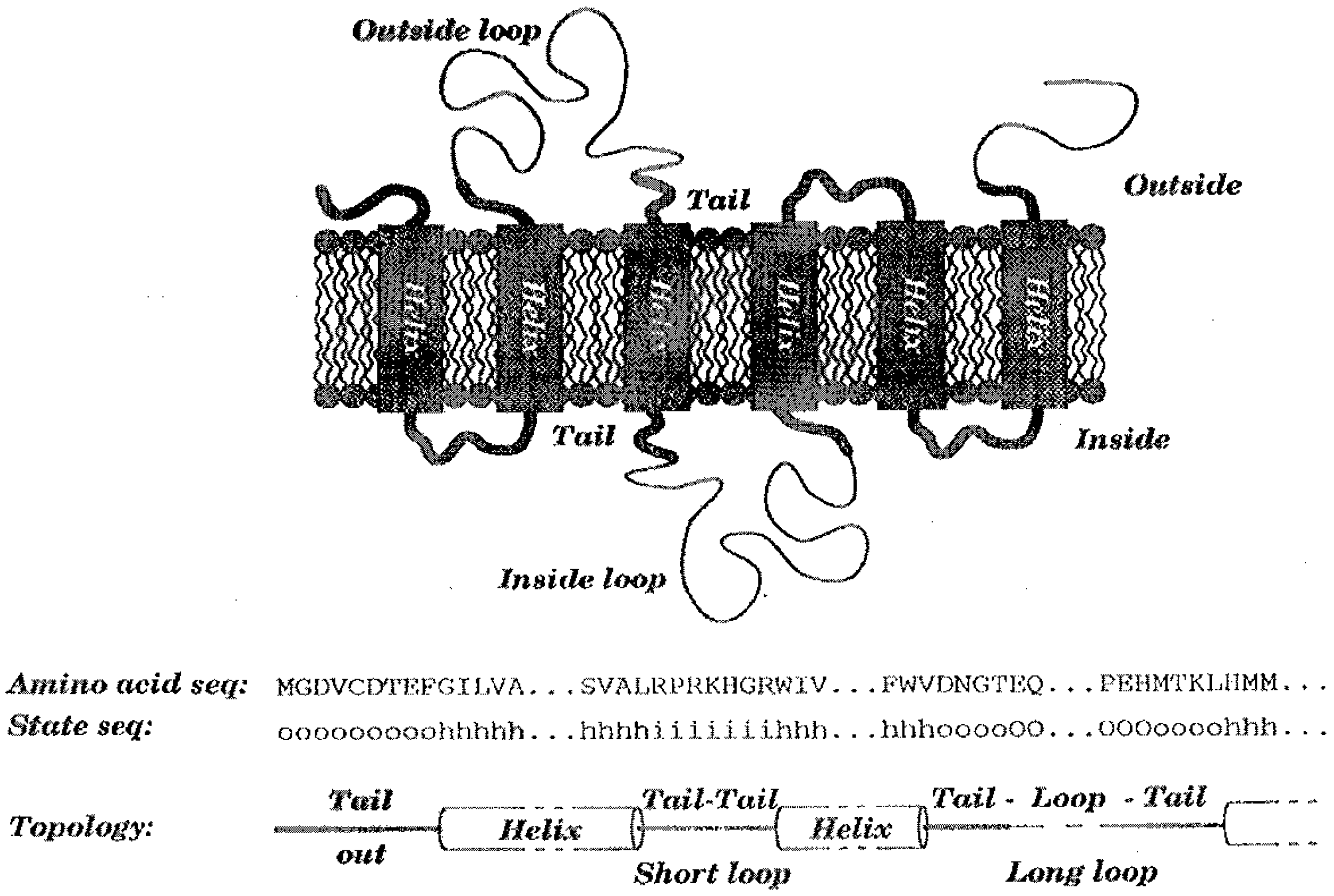


Fig. 1. DAS plot of two arbitrarily chosen proteins (COX3_PARDE versus CYDB_ECOLI). The cross weighted cumulative score profile (dotted line) and the global DAS profile (continuous line) calculated as the average of the cumulative score profiles obtained for comparisons with the other 43 proteins in the test set are also shown for COX3_PARDE. COX3_PARDE has seven and CYDB_ECOLI has eight transmembrane segments.

$$F = \sum_{i=1}^{5} \sum_{j=1}^{20} q_{ij} \cdot log \left(\frac{q_{ij}}{p_j} \right)$$

- where p_j is the frequency of the residue j in the whole protein;
- q_{ij} is the frequency of the residue j in the structural part i of the protein.



Amino acid seq: `MGDVCDTEFGILVA...SVALRPRKHGRWIV...FWVDNGTEQ...PEHMTKLHMM...`

State seq: `ooooooooohhhhh...hhhhiiiiiiihhh...hhhooooOO...OOOoooohhh...`

Figure 2. Structural states defined for a typical helical transmembrane protein. The five states are: inside loop (I), inside tail (i), membrane helix (h), outside tail (o) and outside loop (O). Tails (thick lines) are thought to interact with the inside or outside parts of the membrane, while loops (thin lines) do not. Two tails between helices can form a short loop, but longer loops are formed by tail-loop-tail sequences.

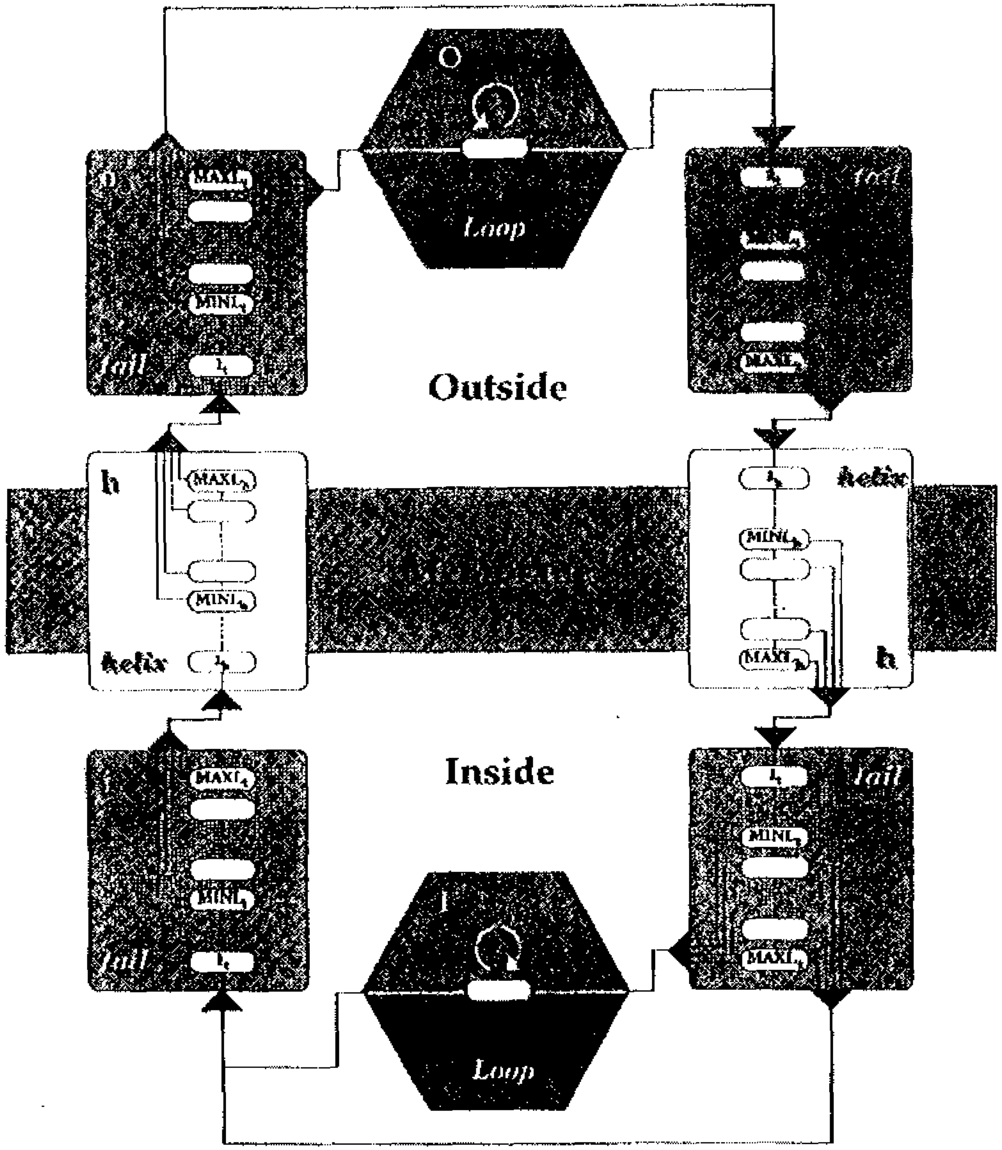


Figure 3. Architecture of HMM used for topology prediction. States with the same transition matrices are colored in the same way: white, helix states; light gray, tail states; dark gray, loop states. Rectangular areas FL type states; hexagonal ones, NFL type states. The observation-symbol probabilities used by states are marked in each state. The structure of substates in the case of the FL type is drawn within states. Lines and arrows show the possible transition between states or substates.

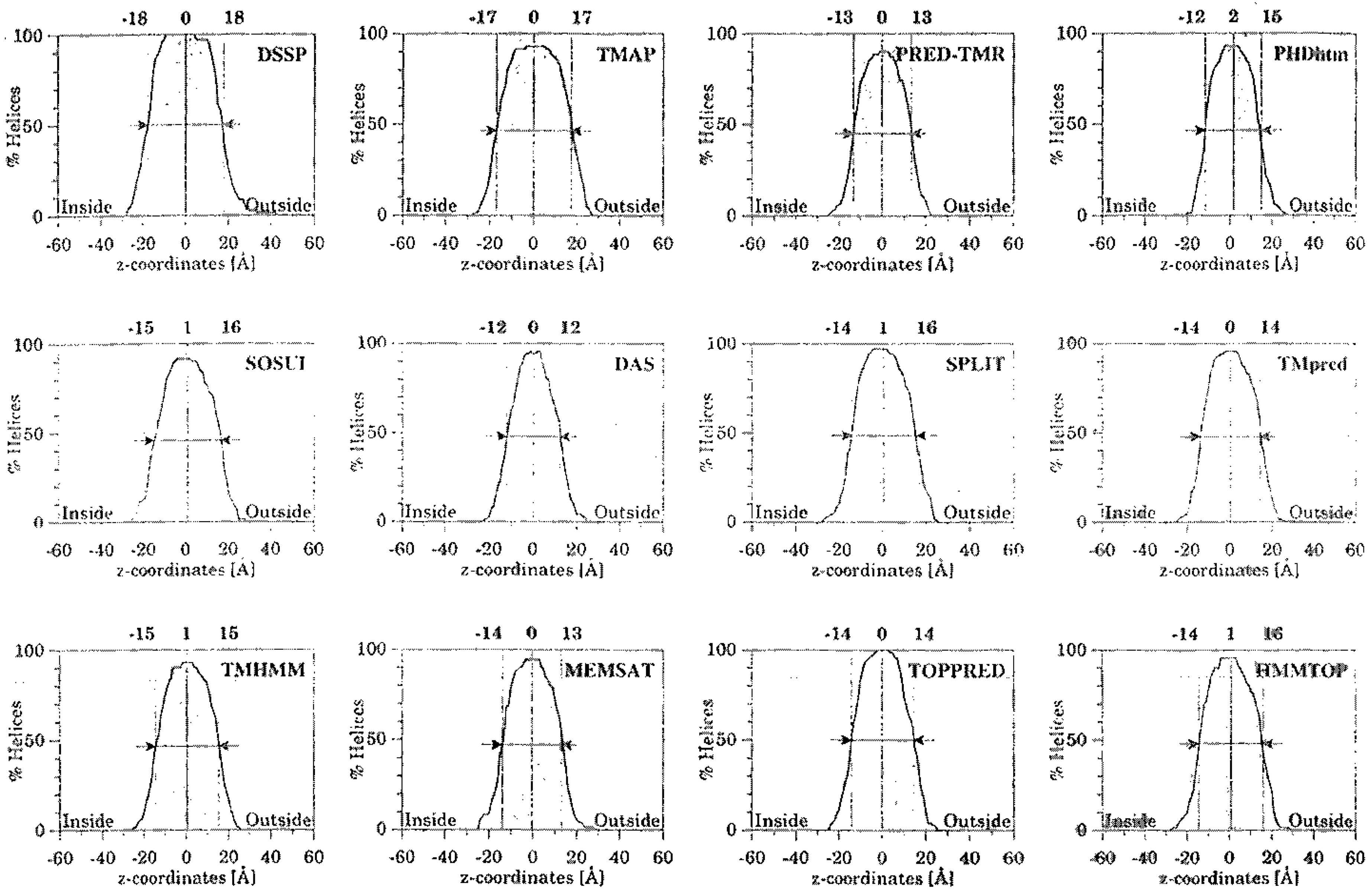


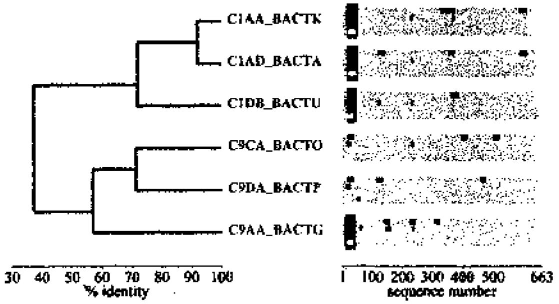
Figure 2. Distribution of transmembrane helices measured parallel to the average direction of the transmembrane helices in 1 Å slices in the eight selected proteins (see the text). Transmembrane helices are defined by the DSSP program and predicted by various transmembrane helix prediction methods. The *z*-coordinates at the value of 50% and at the mean of the curves are shown above the graphs. The gray areas show the same regions as in Figure 1.

Prediction method	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP
Number of TMHs	**0**				**1**				**2**				**3**				**more than 3**			
Globular	523				0				0				0				0			
proteins	269	394	475	453	158	107	35	56	62	17	13	12	21	5	0	2	13	0	0	0
Transmembrane	0				43				7				6				102			
proteins	0	0	1	1	25	30	42	39	12	9	6	10	13	17	10	7	108	102	99	101
Prion proteins	0	0	0	0	2	2	2	2	0	2	4	0	4	2	0	4	0	0	0	0

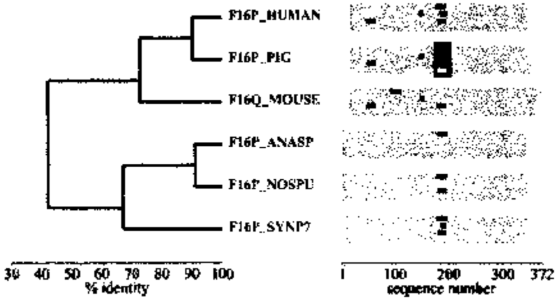
Prediction methods	Globular (total=523)	Transmembrane (total=158)	Sum (total=681)
1	247 (47%)	158 (100%)	405 (59%)
2	455 (87%)	158 (100%)	613 (90%)
3	512 (98%)	157 (99%)	669 (98%)
4	522 (99.8%)	156 (99%)	678 (99.6%)

Globular proteins

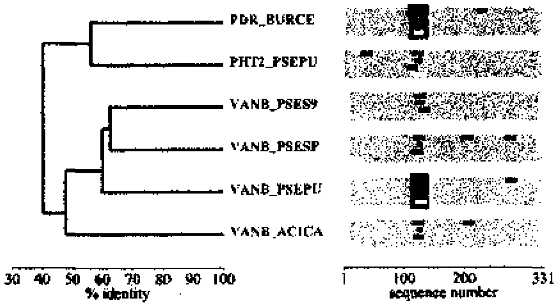
Insecticidal toxin (1ciy)



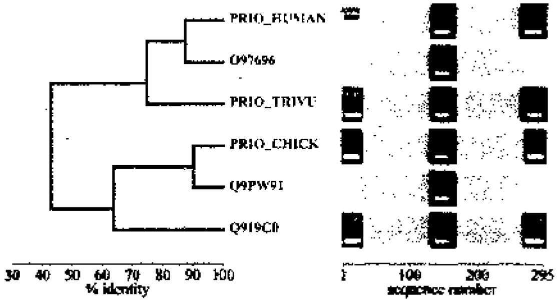
Fructose-1,6-bisphosphatase (1frp)



Phthalate dioxygenase reductase (2pia)

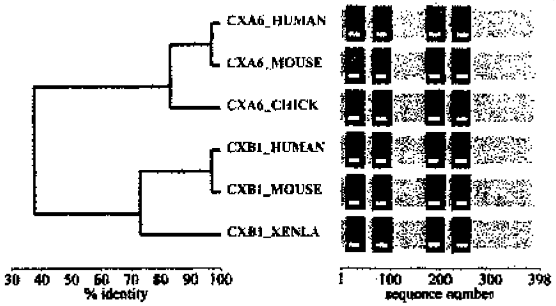


Prion proteins

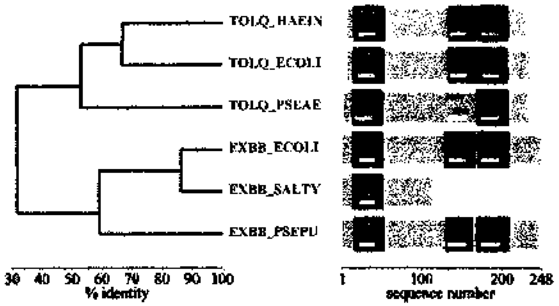


Transmembrane proteins

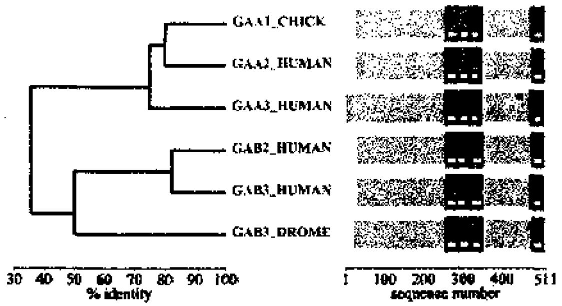
Gap junction β-1 protein

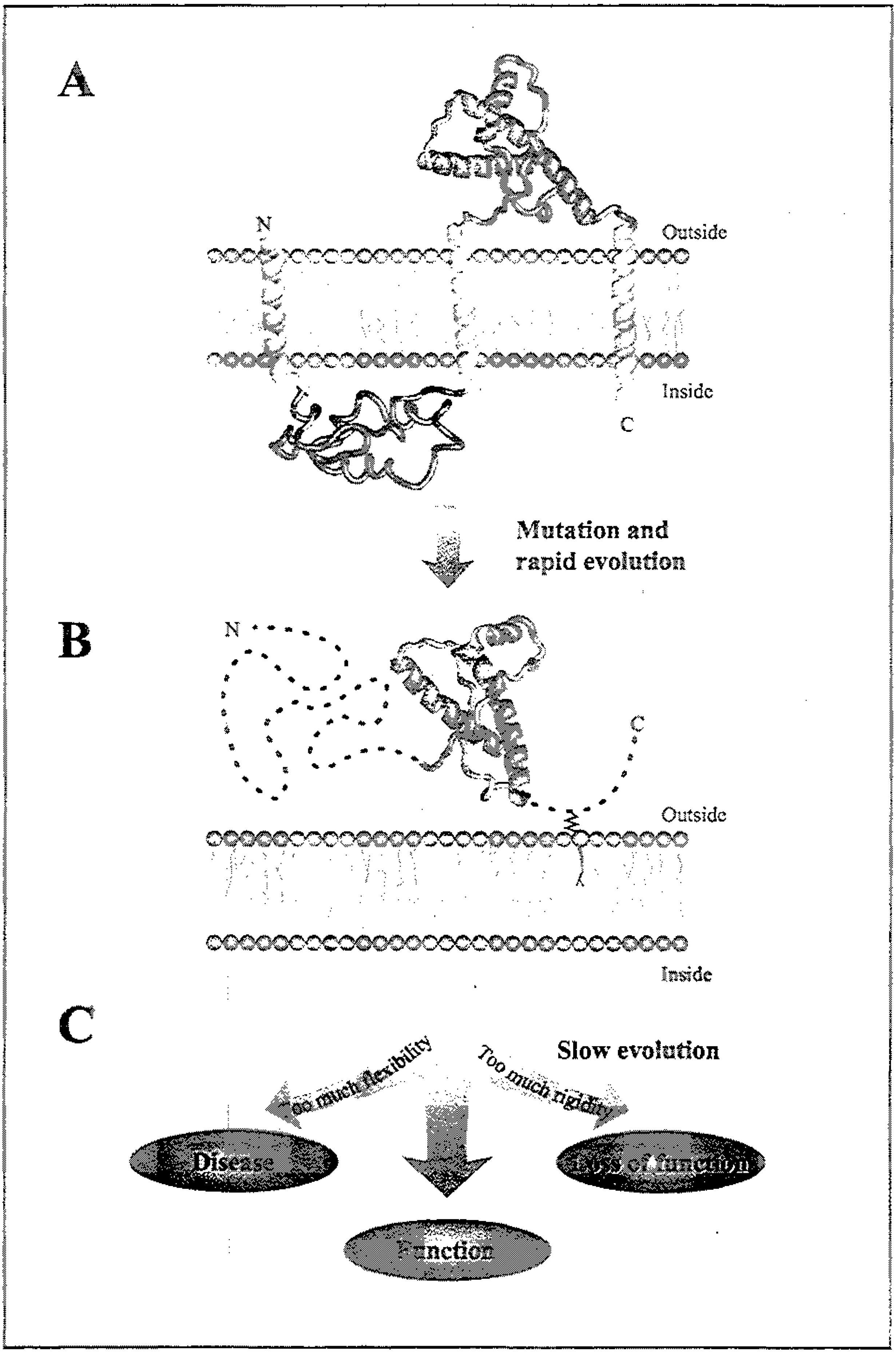


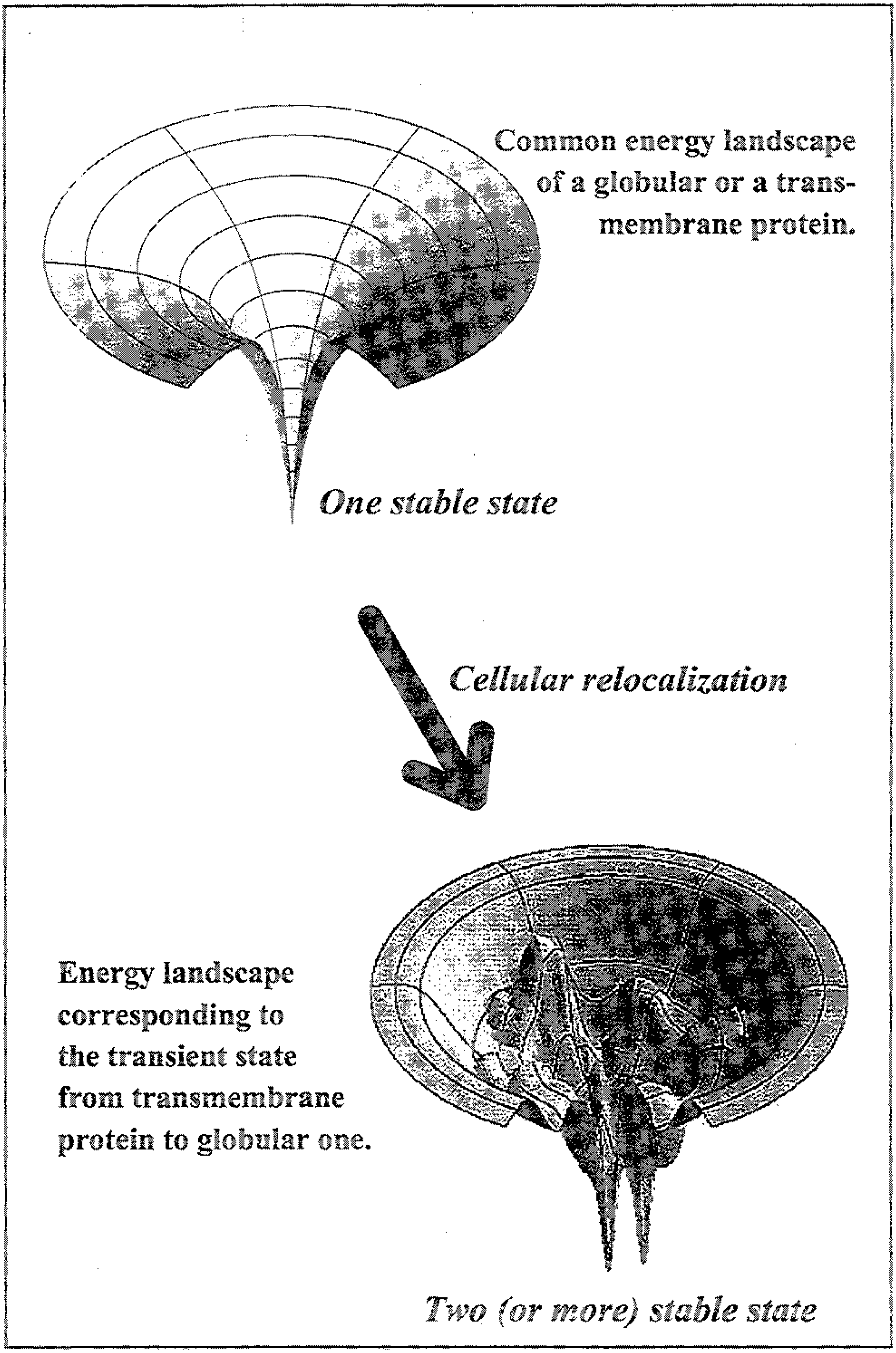
Biopolymer transport protein



Gamma-aminobutyric-acid receptor




A
N
Outside
Inside
C
Mutation and rapid evolution
B
N
C
Outside
Inside
C
Slow evolution
Too much flexibility
Too much rigidity
Disease
Loss of function
Function


Common energy landscape of a globular or a trans-membrane protein.
One stable state
Cellular relocalization
Energy landscape corresponding to the transient state from transmembrane protein to globular one.
Two (or more) stable state