Interaction-Relation Based Computation Approaches for Predicting Interaction Sites of Cytochrome and Photosystem I

J. Wisniewski, W. Chen, A. Sekmen, * B. Bruce

Computer Science, Tennessee State University, Nashville, TN

* Biochemistry, Cellular and Molecular Biology, University of Tennessee at Knoxville

Research Purpose

GOAL The research addresses the problem of computationally predicting the interaction sites of protein pairs (donors and acceptors) that tap into photosynthetic processes to efficiently produce inexpensive hydrogen.

MOTIVATION

Hydrogen is particularly useful energy carrier for transportation. However, there are no sources of molecular hydrogen on the planet. An attractive solar based approach is bio-hydrogen production, which utilizes protein components, Photosystem I (PSI) and cytochrome c6 (cry c6).

Due to the lack of a crystal structure for bound binary complexes, traditional structural biology tools are rendered unavailable to date. We focus on computational approaches that can analyze the protein sequences from these systems and predict the interacting residues of cyt c6 and PSI protein pair.

Why Computationally Predicting Interaction

Biologist's Approach

✓Precise

✓Expensive and time consuming

Example:

Using mass spectrometry to find the interaction sites of 1 pair of proteins from c6 and PsaF (more than 1 day?)

Computer Scientist's Approach

✓Predict best candidates for the Biologist

✓Inexpensive and time efficient

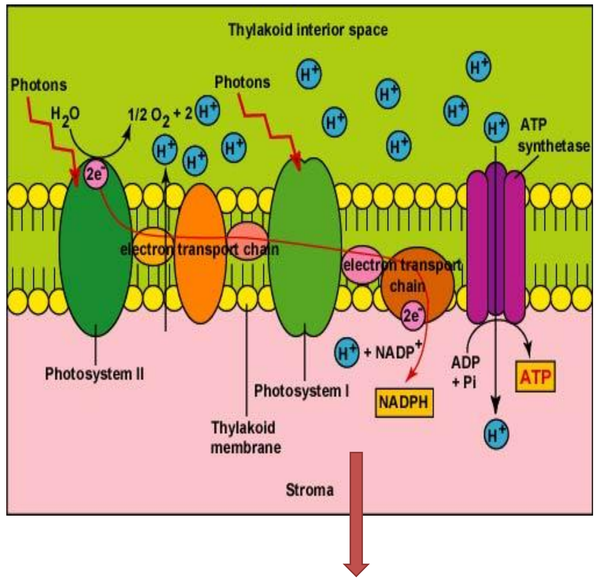
Example:

Runinng a computing program to calculate the possible interaction sites of 400 pairs of proteins (5 munites?)

Acknowledgement

This work is supported by a collaborative contract from NSF HBCU-UP and TN-SCORE (NSF EPS 1004083 and NSF 1137484)

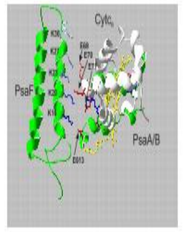
Natural Photosynthetic Process to Produce Hydrogen

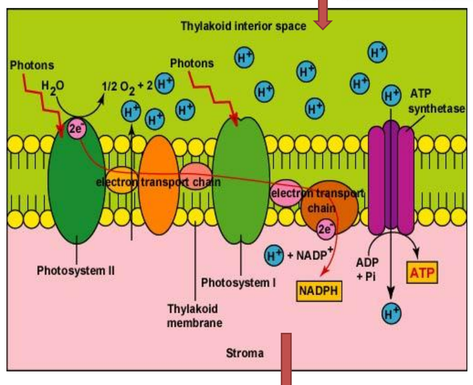


Natural Photosynthetic Process is not efficient and quantitative !!!

Artificially Photosynthetic Process to Produce Hydrogen

Artificially add the proteins that can donate and accept large number of electrons **at protein interaction sites to produce large quantity of energy**





Finding the interaction sites of proteins is critical for developing the system that tap into photosynthetic processes to efficiently produce inexpensive energy !!!

Interaction Relation

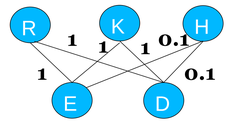
Interaction relation is built on amino acid bonds

Three Types of Interaction

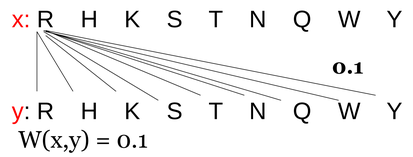
1. Electrostatic Bonding
2. Hydrogen Bonding
3. Hydrophobic Bonding

Weight Scheme

Wight for electrostatic bond

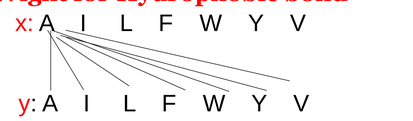


Wight for Hydrogen bond



W(x,y) = 0.1

Wight for Hydrophobic bond



A = 3.1, I = 5.8, L = 5.1, F = 4.1, W = 0.4, Y = 0, V = 5.5; W(x,y) = (W(x) + W(y])/11.6 * 0.1

Algorithms

Three Prediction Models:

1. *Electrostatic bond Only*
2. *Electrostatic bond + Hydrogen bond*
3. *Electrostatic bond + Hydrogen bond + Hydrophobic bond*

Approaches:

1. *Using Dynamic Programming Technique to calculate the interaction score for all pairs of subsequences in PSI and C6*
2. *Using window or using gap*

Window	**Gap**
******R******	******R******
EE*R***	***EE*R******

Interaction Score Calculation: (for window only)

$$S[i,j] = \begin{cases} 0, & \text{if } i = 0 \text{ or } j = 0 \\ \max\{S[i\text{-}1,j\text{-}1] + W(x_i, y_j),\ 0\}, & \text{Otherwise} \end{cases}$$

Example of Interaction Score Calculation (window)

S		1	2	3	4	5	6	7	8	9
		A	E	L	M	D	S	E	A	E
	0	0	0	0	0	0	0	0	0	0
1	G	0	0	0.2	0	0	0.2	0	0.2	0	0.2
2	P	0	0	0.2	0	0	0.2	0	0.2	0	0.2
3	R	0	0.2	1	0.4	0.2	1	0.4	1	0.4	1
4	F	0	0	0.4	0.78	0.18	0.4	0.78	0.6	0.78	0.6
5	K	0	0.2	1	0.6	0.98	1.18	0.6	1.78	0.8	1.78
6	Y	0	0	0.4	0.78	0.38	1.18	0.96	0.8	1.56	1
7	K	0	0.2	1	0.6	0.98	1.38	1.38	1.06	1	2.56
8	H	0	0.02	0.3	1.02	0.62	1.08	1.4	1.48	1.98	1.1

Results

Results of predicting 86 pairs of psaF and C6 sequences in same organism

First pair of PsaF and C6:

Psaf:MRRLFALILAIGLWFNFAPQAQALGANLVPCKDSPAFQALAENARNT
TADPESGKKRFDRYSQALCGPEGYPHLIVDGRLDRAGDFLIPSILFLYIAGW
IGWVGRAYLQAIKKESDTEQKEIQIDLGLALPIISTGFAWPAAAIKELLSGEL
TAKDSEIPISPR

c6:MENVGCEENLLRLILVNLLLVIALLCNLTIIYPALAAETSNGSKIFNANCA
ACHIGGANILVEHKTLQKSGLSKYLENYEIEPIQAIINQIQNGKSAMPAFKN
KLSEQEILEVTAYIFQKAETGW

Three interaction sites with top three scores:

1st interaction site information:
Interaction score: 2.76
Interaction site location and subsequence in Psaf:
54-59, u = KKRFDR
Interaction site and subsequence in c6:
106-111, v = EQEILE
.....

Statistics for 86 PsaF and c6 sequences