

N-end rule 기반 단백질 수명 예측 프로그램의 한계

Limitation of Protein Half-life Prediction in Computational Approach Based on N-end Rule

류 제 운, 김 학 용

충북대학교 생화학과

Ryu, Jea Woon and Kim, Hak Yong

Chungbuk National Univ.

요약

본 연구는 N말단 아미노산 종류에 따라 단백질의 수명(half-life)이 결정된다는 N-end rule을 기반으로 단백질 수명을 예측해주는 프로그램인 protparam의 결과와 bleach-chase를 이용한 실험 데이터를 비교·분석하였다. 단백질 수명을 결정하는 여러 요인들을 고려하지 않고 한 가지 요인만을 반영한 protparam의 결과는 실제 측정값과 현격한 차이를 나타낸다. 특히 실제 단백질은 NME (N-terminal Methionine Excision) 현상이 일어나는데 이를 고려하지 않고 유전체에서 번역한 그대로의 아미노산 서열을 가지고 단백질 수명을 계산하는 한계를 가지고 있다. 이에 본 연구에서 N말단 아미노산을 순차적으로 제거하여 N-end rule을 적용한 결과도 실험 데이터와 일치하지 않는 결과를 보여주고 있음을 확인하였다. 따라서 현재 사용되고 있는 단백질 수명 예측 프로그램은 이런 문제점을 가지고 있기 때문에 새로운 예측 알고리즘의 개발이 요구된다.

I. 서론

생물체 네트워크의 완벽한 이해는 전사, 번역, 수식뿐만 아니라 분해(degradation)에 대한 이해가 요구된다. 하지만 대량(high-throughput)으로 단백질의 조절관련 시스템의 접근방법의 한계로 많이 연구가 되지 못한 것이 사실이다. 신약 및 신호전이 경로에서 단백질의 분해는 조절에 있어 중요한 역할을 하게 된다.

1986년 N-말단 아미노산 종류에 따라 단백질의 수명(half-life)이 결정된다는 N-end rule은 비교적 근사한 수명예측 방법이며 또한 실험에 의해 N-말단의 아미노산이 밝혀진 경우에는 적용이 가능하나, 컴퓨터를 이용한 예측에 직접 적용하기 위해서 NME 현상을 고려하여야 할 것이다 [1].

단백질은 합성된 후 signal peptide에 따라 sorting 과정을 거쳐 소기관으로 이동되어 그 역할을 수행한다. 일반적으로 단백질은 소기관에 도착한 후 signal peptide가 제거된다. 대부분의 단백질은 signal peptide가 N-말단에 존재하기 때문에 N-end rule에 의한 단백질 수명 예측에 signal peptide를 고려하여야 한다.

또한 단백질 수명에 중요한 역할을 하는 UFD (ubiquitin-fusion degradation) 경로를 고려하지 않은 알고리즘의 사용은 한계가 있을 것이다.

단백질 수명에 결정적인 역할을 하는 이러한 세 가지 인자를 고려하지 않고 단백질의 수명을 예측하는 알고리즘의 한계를 분석하고자 우리는 NME 현상과 signal peptide 제거 현상을 고려한 단백질 수명 계산값과 실험값을 비교하였다. 그러나 UFD 값을 고려한 예측은 충분

한 데이터 부족으로 분석하지 못하였다.

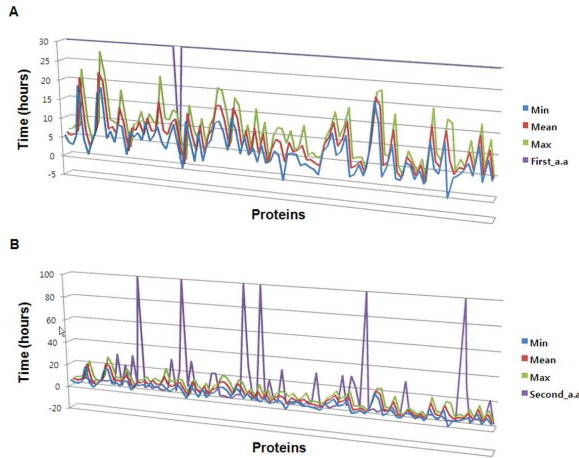
II. 본론

1. Dataset

Fluorescently tagged proteins를 이용한 bleach-chase 기법을 통해 living human cells의 100개 단백질 실험 데이터를 얻어 분석하였다 [2]. ExPASy proteomics Server에서 제공하는 ProtParam tool을 이용하여 단백질 수명 데이터를 얻어 분석하였다 [3].

2. Result

각 단백질의 fasta 서열을 uniprot을 통해 얻고 각 N-말단의 아미노산을 기준으로 protparam을 통해 수명을 계산한 결과 실제 실험값과 많은 차이가 발생함을 알았다 (그림 1A). 이는 실제 생체 내에서 단백질은 NME 현상이 일어나는데 이를 고려하지 않고 유전체에서 번역한 그대로의 아미노산 서열을 가지고 단백질의 수명을 계산하기 때문이다. 이에 우리는 각 단백질의 fasta 서열의 N-말단을 기준으로 10번째 아미노산까지 순차적으로 제거하였다. 각각의 서열을 protparam을 적용한 단백질 수명을 계산한 값과 실제 실험값을 비교해 보았다 (그림 1B, 표 1).



▶▶ 그림 1. protparam 결과와 실험값과의 비교

순차적으로 제거한 각 아미노산 서열의 계산한 수명값과 실험값을 비교하였을 때 약 10% 전후로 나타났는데, 이는 protoparam 프로그램을 이용하여 단백질의 수명을 예측하는 방법에 한계가 있음을 의미 한다 (표 1).

실험값이 있는 100개의 단백질 중에서 이 signal peptide 정보를 가지고 있는 단백질이 7로 그 수가 매우 적었다. 이 적은 데이터로 N-end rule을 적용하여 단백질 수명 예측의 한계를 제시하기에 한계가 있지만, 7개 단백질 중 단 한 개만이 실험값과 일치함을 볼 수 있었다 (표 2).

표 1. 순차적 아미노산 제거시 실험값과 계산값의 일치도 분포

Position	%	Protein List	a.a
1	0	-	-
2	13	BAG1,C1QBP,COPS6,HDAC2,PSMA3,PSMB6,PSMB7,PSMD12,RPL29,RPS3,STMN1,TUBA1C,UBE2V2	A,L,R,S
3	10	CKS2,COPS6,COTL1,GAPDH,HDAC2,POMP,PSMA3,PSMB6,PSMB7,TUBA1C	A,E,H,K,S,T,Y
4	7	C1QBP,CKS2,COPS6,ENO1,PSMB6,RPL39,TUBA1C	A,C,H,K,L,T
5	7	C1QBP,COPS6,DDX18,DNMT1,GAPDH,PSMB6,RPL27A	A,K,L,P,T
6	8	CALM1,COPS6,COX7C,POMP,PSMA3,PSMB6,RPS7,TUBA1C	A,I,L,S,T
7	8	BAG1,CKS2,COPS6,NDUFAF2,PSMB6,PSMC4,RPL29,TUBB2C	A,H,L,Y
8	9	CIRBP,CKS2,COPS6,PSMA3,PSMA7,PSMB6,PSMB7,RPS8,TUBA1C	A,H,L,T,W,Y
9	6	CKS2,COPS6,GAPDH,PSMA3,PSMD12,TXN	A,D,N,S,T
10	7	CKS2,COPS6,EIF4EBP1,POMP,PSMA3,RPL29,RPS3A	A,D,H,L,T

* a : amino acids

표 2. signal peptide의 위치와 실험값과 계산값의 비교

Protein	Start	End	a.a	Protparam	Min	Max	+/-
C1QBP	1	73	L	5.5	3.2	8.2	+
CD44	1	20	Q	0.8	18.9	22.7	-
COX7C	1	16	S	1.9	17.4	27.6	-
NDUFB11	1	29	S	1.9	5.1	5.3	-
PRDX5	1	52	M	30	6.8	12	-
SSBP1	1	16	E	1	6.6	7.2	-
TFAM	1	42	S	1.9	12.6	20	-

* +/- : +는 실험값과 계산값이 일치하고 -는 불일치함을 나타낸다.

III. 결론

본 연구에서는 단백질의 수명을 계산하기 위해 사용되는 protparam의 결과 값을 실험값과 비교하여 컴퓨터 프로그램에 의한 단백질 수명 예측의 타당성을 분석하였다. 단백질 수명을 결정하는 NME, signal peptide, UFD에 대한 충분한 데이터가 존재하고 이를 고려한 수명 예측 프로그램이 필요한데, 현 시점에서 이러한 데이터 부족과 N-end rule만을 고려한 단백질 예측 프로그램은 한계를 가지고 있다. 따라서 본 연구진을 향후 NME, signal peptide, 및 UFD를 고려한 단백질 예측 프로그램을 개발하고자 한다.

■ 참고 문헌 ■

[1] Bachmair A., Finley D., Varshavsky A., "In vivo half-life of a protein is a function of its amino-terminal residue", Science, Vol, 234, no. 4773, pp. 179-86, 1986.

[2] Eden E., Geva-Zatorsky N., Issaeva I., Cohen A., Dekel E., Danon T., Cohen L., Mayo A., Alon U., "proteome half-life dynamics in living human cells", Science, Vol, 331, No. 6018, pp. 764-768, 2011.

[3] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., "Protein Identification and Analysis Tools on the ExPASy Server", The Proteomics Protocols Handbook, Humana Press, pp. 571-607, 2005.

* 본 논문은 교육과학기술부의 "지역거점연구단육성사업/충북 BIT연구중심대학사업단"의 지원을 받아 수행되었음.