

총 설

단백질 폴딩

송실대학교 생명정보학과
신향철

단백질은 생명체를 구성하는 가장 중요한 물질로써 생명체를 이루는 세포의 골격형성, 물질대사, 유전물질의 합성, 세포내 신호전달 및 반응 등 생명체가 의존하는 모든 현상을 조절하는데 관여하고 있다. 이러한 단백질의 기능은 개개의 단백질이 갖고 있는 독특한 3차구조에 의해 나타나는데 이들 단백질이 리보솜(ribosome)에서 생합성될때는 단순히 아미노산들이 펩타이드 결합에 의해 연결된 1차원적인 사슬 형태로 되어있다. 따라서 활성을 갖기 위해서는 독특한 3차구조로 바뀌어야 하는데 이 과정을 단백질 폴딩(protein folding)이라고 하며 아직까지 해결이 안된 생명과학분야의 숙제로 남아있다.

최근 생물학의 발전은 인간을 포함한 여러 동식물들의 유전체의 염기서열을 결정하는 게놈과제(genome project)를 이미 끝냈거나 또는 수행 중에 있으며, 이와 더불어 게놈 후 연구의 중심축이 생체내 유전자들이 코우딩하고 있는 모든 단백질들의 기능 규명연구에 초점이 맞추어지고 있다. 이미 단백질-단백질 interaction network 규명연구, 단백질 폴딩(fold) 또는 기능의 비교를 통한 단백질 진화연구 등 규모가 큰 연구과제들이 진행되고 있으며, 이러한 노력은 결국 생명체의 본질을 이해하고 나아가 질병의 발생기작 이해 및 이에 대한 치료법 개발 등 인류의 복지에 크게 이바지 할 것으로 보여진다. 이러한 모든 연구의 바탕에는 단백질의 구조-기능 상관관계가 자리잡고 있으며, 어떻게 유전자가 코우딩하는 아미노산 서열이 단백질의 고유한 3차구조에 대한 정보를 갖고 있는가를 이해하여야 하는 근원적인 문제를 내포하고 있다.

단백질폴딩은 아직 구조형성이 안된 1차원적인 아미노산 사슬이 구조변형을 통해 독특한 3차구조를 갖는 과정을 말한다. 따라서 폴딩문제는 왜 그리고 어떻게 이러한 과정이 일어나는지를 결정하는 일이며, 폴딩과정 동안에 일어나는 구조변화와 이러한 변화를 지배하는 작용을 해석하는 일이다. 폴딩연구의 중요한 목표는 아미노산 서열, chain topology, pH, 염 농도, 온도 등과 같은 인자들이 폴딩과정의 속도와 열역학에 미치는 효과에 대한 정량적인 예측을 가능케 하는 모델을 개발하는데 있다. 그러한 모델의 개발은 폴딩과정 중에 나타나는 여러 다른 폴딩 중간체들의 구조적, 열역학적 특성과 이들 중간체들 사이에서의 변화 속도에 대한 자세한 지식이 요구된다.

폴딩 모델들

단백질 폴딩연구는 영국의 Sanger가[1] 인슐린의 아미노산 서열을 밝힘으로써 아미노산 서열은 단백질마다 고유하다는 사실을 증명한 후 얼마 되지 않아 denaturation/renaturation(unfolding/refolding) 실험을 통해 시작되었다. Ribonuclease A를 모델 단백질로 이용하여 unfolded된 폴리펩타이드 사슬이 활성을 갖는 native 단백질로 자발적으로 refolding 될 수 있음을 보임으로써 Anfinsen은 단백질의 3차구조에 필요한 모든 정보는 아미노산 서열에 수록되어 있다고 결론지었다[2]. 이후의 연구는 서론에서도 언급한 왜 그리고 어떻게 유전자에 수록된 단백질의 1차원적인 아미노산 서열(genotype)이 고유한 3차구조로 표현(phenotype)되는가를 밝히는데 집중되어 왔다.

1960년대에 Columbia 대학에 재직중이던 Levinthal은 간단한 생각전환을 통해 이후의 폴딩연구에 방향을 제시하는 획기적인 업적을 이루어 냈다[3]. 간단히 설명하면 다음과 같다. 예를 들어 100개의 아미노산으로 이루어진 단백질이 있는데, 각각의 아미노산이 helix, sheet, 또는 random 구조중의 한 구조를 갖는다고 가정하면 (실제로는 이보다 훨씬 많은 구조가 가능하나 여기서는 간단히 했음), 이 사슬이 가질 수 있는 구조의 수는 3^{100} (or about 10^{48}) 이다. 만일 각각의 아미노산이 한 구조에서 다른 구조로 변화하는데 10^{14} s^{-1} (작은 분자들에 있어서의 bond rotation 속도)이 걸린다면, random search를 통해 native 구조를 찾는 데 걸리는 시간은 10^{34} 초 또는 10^{26} 년이 걸린다는 계산이 나오는데 이는 우주의 나이보다도 더 긴 시간이다. 실제의 경우 100개 정도의 아미노산으로 이루어진 단백질이 폴딩하는데 걸리는 시간은 수백 millisecond 에서 수분정도가 걸린다. 이러한 계산상의 예측과 실제 폴딩사이의 차이를 Levinthal paradox라고 하며, 이러한 모순을 해결하기 위해서는 폴딩경로가 존재해야 한다는 사실을 제시하였다.

1970년대에 폴딩 기작을 설명하기 위한 3가지 모델이 태동하였는데 이들 모델은 2차구조 형성과 3차구조 형성이 분리되어 나타나는 것을 가정하여 제시한 모델들이다 (Figure 1). Framework model (or diffusion-collision-adhesion model) 은 native한 2차구조들이 먼

저 형성되고 이들 2차구조가 diffusion-collision 기작에 의해 native한 3차 구조를 형성하는 것으로 가정하고 있다 [4,5]. Nucleation/growth model은 서로 인접한 아미노산들이 하나의 native 구조를 형성하고 이를 핵으로 하여 순차적으로 native한 구조가 만들어 진다고 가정하였다 [6]. 따라서 3차구조는 2차구조의 결과로 만들어지게 된다. Hydrophobic collapse model은 폴딩 초기단계에 소수성(hydrophobic) side chain 들이 용매인 물로부터 감추어지며(hydrophobic collapse), molten globule 이라고 불리는 중간단계를 형성하고 이후 재배열(rearrangement)을 통해 native 3차구조를 형성한다고 가정한다 [7,8]. 여기서는 3차구조형성 후에 2차구조가 만들어진다. 이들 모델과는 전혀 다른 jigsaw puzzle model이 있는데, 이 model에서는 마치 jigsaw puzzle 게임에서처럼 폴딩 경로가 하나가 아니고 여러 다른 경로를 통해 다양하게 나타날 수 있다고 가정하였다 [9]. 이후에 여러 실험 및 계산에서 나타난 결과들을 통해 이들 모델들을 상당부분 포괄하는 새로운 견해가 나타나게 된다.

기기의 발전이 과학의 진보에 얼마나 중요한가는 새삼스러운 일이 아니지만 폴딩 연구에 있어서도 비슷한 현상이 나타났다. 앞서의 모델들이 나온 이후 기기의 발전은 수 nanosecond에서 수 millisecond 사이에서 일어나는 현상을 탐색할 수 있게 되었으며(Table 1 참조), 이러한 연구를 통해 나타난 연구결과들은 폴딩현상을 보다 더 깊이있게 이해하는데 많은 도움을 주게 되었다.

최근의 폴딩에 대한 견해는 개개의 단백질에 대해 하나의 고유한 폴딩경로가 존재하는 것이 아니라 수없이 많은 폴딩경로가 존재할 수 있으며, 이는 마치 funnel을 따라 물이 흘러내려가는 것과 같은 원리로 설명할 수 있다고 본다(Figure 2 참조). 물방울 하나가 funnel을 따라 내려가는 길은 무수히 많이 존재할 수 있으며, 특히 맨 위의 출발하는 지점(Figure 2는 2차 원으로 그렸으나 3차원 모양에서는 funnel의 맨 위의 둥굴기 때문에 테두리의 어느 한 점을 말함) 그 이후의 내려가는 경로에 주된 영향을 주게 된다. 여러 단백질에서 multiple pathway가 발견되어 이러한 견해를 뒷받침하고 있으며[10-13], 과거에는 불가능했던 unfolded된 상태의 구조에 대한 접근이 NMR의 발전과 더불어 점차 가능해 지고 있어 앞으로 초기단계를 연구하는데 크게 공헌할 것으로 보인다[14]. 따라서 unfolded된 polypeptide가 어떻게 폴딩을 시작하는가 하는 초기단계의 폴딩문제가 중요한 사안의 하나로 등장하였으며, unfolded된 상태에서는 단백질 분자들이 서로 다른 상태(빠르게 변화하는 local 구조들을 가짐)의 ensemble로 존재하는데 이러한 multiple starting points에서 어떤

게 1개의 native 구조로 converge하는가를 규명하는 문제로 초점이 맞추어지고 있다.

세포내에서의 폴딩

세포내의 단백질 농도는 상당히 높으며 (~300g/L의 농도) 상대적으로 높은 생체온도(~37°C)로 인해 분자운동이 활발하여 중간체들간에 또는 중간체와 다른 단백질 간에 aggregates가 쉽게 생길 수 있다. 이러한 문제를 해결하기 위해 생명체들은 폴딩을 도와주는 helper 단백질들을 진화시켜 왔는데 이들을 통칭 molecular chaperone이라 한다. 'molecular chaperone'이라는 term은 DNA와 histone이 nucleosome으로 assembly하는 것을 도와주는 단백질인 nucleoplasmin의 역할을 설명하기 위해서 쓰이기 시작하였는데 [15], 현재까지 약 20여 가지가 알려져 있다. 잘 알려진 molecular chaperone으로 hsp70과 hsp60이 있는데 hsp70에 속하는 단백질로는 원핵생물세포의 DnaK와 진핵생물세포의 Hsc/Hsp70, Ssa 또는 Ssb 등이 알려져 있는데, hsp70은 리보솜에서 단백질 생합성시 약 7개정도의 hydrophobic한 아미노산 서열과 결합하여 생합성 도중에 aggregation이 일어나는 현상을 방지하는 역할을 한다. Hsp60에 속하는 단백질로는 원핵생물세포의 GroEL과 진핵생물세포의 TriC/CCT 등이 알려져 있는데, GroEL은 여러 group에서의 연구를 통해 상당히 잘 알려져 있다 [16-18]. GroEL은 두개의 도너츠를 쌓아 놓은 것 같은 구조를 갖고 있는데 생합성이 끝난 단백질의 중간체와 결합하여 aggregation이 일어나는 현상을 막아준다. 이외에 hsp60와 함께 작용하는 chaperone인 hsp10 단백질로 원핵생물세포의 GroES가 알려져 있다. 단백질폴딩에서 proline의 cis-trans isomerization이나, disulfide bond의 형성과 rearrangement 등은 단백질폴딩 속도를 늦추는 역할을 하며 대부분의 폴딩에서 rate-determining step을 형성하는데, 세포내에는 Peptidyl prolyl cis-trans isomerase(PPI)나 Protein disulfide isomerase(PDI) 등이 있어 폴딩을 돕고 있다. 이들 chaperone이 단백질 폴딩기작에 어떤 영향을 미칠 것인가는 많은 관심사였는데 그 동안의 연구결과는 폴딩속도에는 영향을 주지만 폴딩경로나 기작에는 영향을 주지 않는 것으로 나타났다 [19-20]. 따라서 단백질의 폴딩기작과 그 결과로서 나타나는 고유한 3차구조에 대한 정보는 개개 단백질의 아미노산 서열에 있다는 Anfinsen의 주장은 현재 까지 타당한 것으로 받아들여지고 있으며, 아직까지 해결되지 않은 아미노산 서열과 3차구조에 대한 관계식을 알아내기 위한 과학자들의 노력은 계속되고 있다.

REFERENCES

- (1) Sanger, F. (1952) *Adv. Protein Chem.* 7, 1–67.
- (2) Anfinsen, C.B. (1973) *Science* 181, 223–230.
- (3) Levinthal, C. (1968) *J. Chim. Phys.* 65, 44–45.
- (4) Kim, P.S. and Baldwin, R.L. (1990) *Ann. Rev. Biochem.* 59, 631–660.
- (5) Karplus, M. and Weaver, D.L. (1994) *Protein Sci.* 3, 650–668.
- (6) Wetlaufer, D.B. (1990) *Trends Biochem. Sci.* 15, 414–415.
- (7) Ptitsyn, O.B. (1995) *Trends Biochem. Sci.* 20, 376–379.
- (8) Kuwajima, K. (1989) *Proteins: Struct. Funct. Genet.* 6, 87–103.
- (9) Harrison, S.C. and Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4028–4030.
- (10) Rothwarf, D.M., Li, Y.-J. and Scheraga, H.A. (1998) *Biochemistry* 37, 3767–3776.
- (11) van den Berg, B., Chung, E.W., Robinson, C.V., Mateo, P.L. and Dobson, C.M. (1999) *EMBO J.* 18, 4794–4803.
- (12) Creighton, T.E. (1977) *J. Mol. Biol.* 113, 275–293.
- (13) Weissman, J.S. and Kim, P.S. (1991) *Science* 253, 1386–1393.
- (14) Dyson H.J. and Wright P.E. (2001) *Methods Enzymol.* 339, 258–70.
- (15) Laskey, R.A., Honda, B.M., Mills, A.D. and Finch, J.T. (1978) *Nature* 275, 416–420.
- (16) Roseman, A.M., Chen, S.X., White, H., Braig, K. and Saibil, H.R. (1996) *Cell* 87, 241–251.
- (17) Buckle, A.M., Zahn, R. and Fersht, A.R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 3571–3575.
- (18) Agashe, V.R. and Hartl, F.-U. (2000) *Semin. Cell. Dev. Biol.* 11, 15–25.
- (19) Shin, H.-C. and Scheraga, H.A. (2000) *J. Mol. Biol.* 300, 995–1003.
- (20) Coyle, J.E., Texter, F.L., Ashcroft, A.E., Masselos, D., Robinson, C.V., Radford, S.E. (1999) *Nat. Struct. Biol.* 6, 683–90.

Table 1. Experimental techniques used to measure folding

Technique	Timescale	Structural parameter probed
NMR		
(i) Real time	ms-s	Environment of individual residue
(ii) Dynamic NMR	250 μ s	Lineshape analysis provides folding-unfolding rates close to equilibrium
Fluorescence ns-s		
(i) Intrinsic		Environment of Trp and Tyr
(ii) ANS binding		Exposure of hydrophobic surface area
(iii) Substrate binding		Formation of the active site
(iv) FRET		Inter-residue distances
(v) Anisotropy		Correlation time
Circular dichroism ns-s		
(i) Far-UV CD		Secondary structure formation
(ii) Near-UV CD		Tertiary structure formation
FT-IR ns-s Secondary structure formation		
Hydrogen exchange (HX)		
(i) Native stae	min-months	Global stability and metastable states
(ii) Pulsed HX NMR	ms-s	Hydrogen-bond formation in specific residues
(iii) Pulsed HX ESI MS	ms-s	Folding populations
Small-angle X-ray scattering ms		
Atomic force microscopy s		
Unfolding forces and unfolding rate constants of single molecules		
Protein engineering Depends on the method of detection		
Role of individual residues in stabilizing intermediates and transition states		

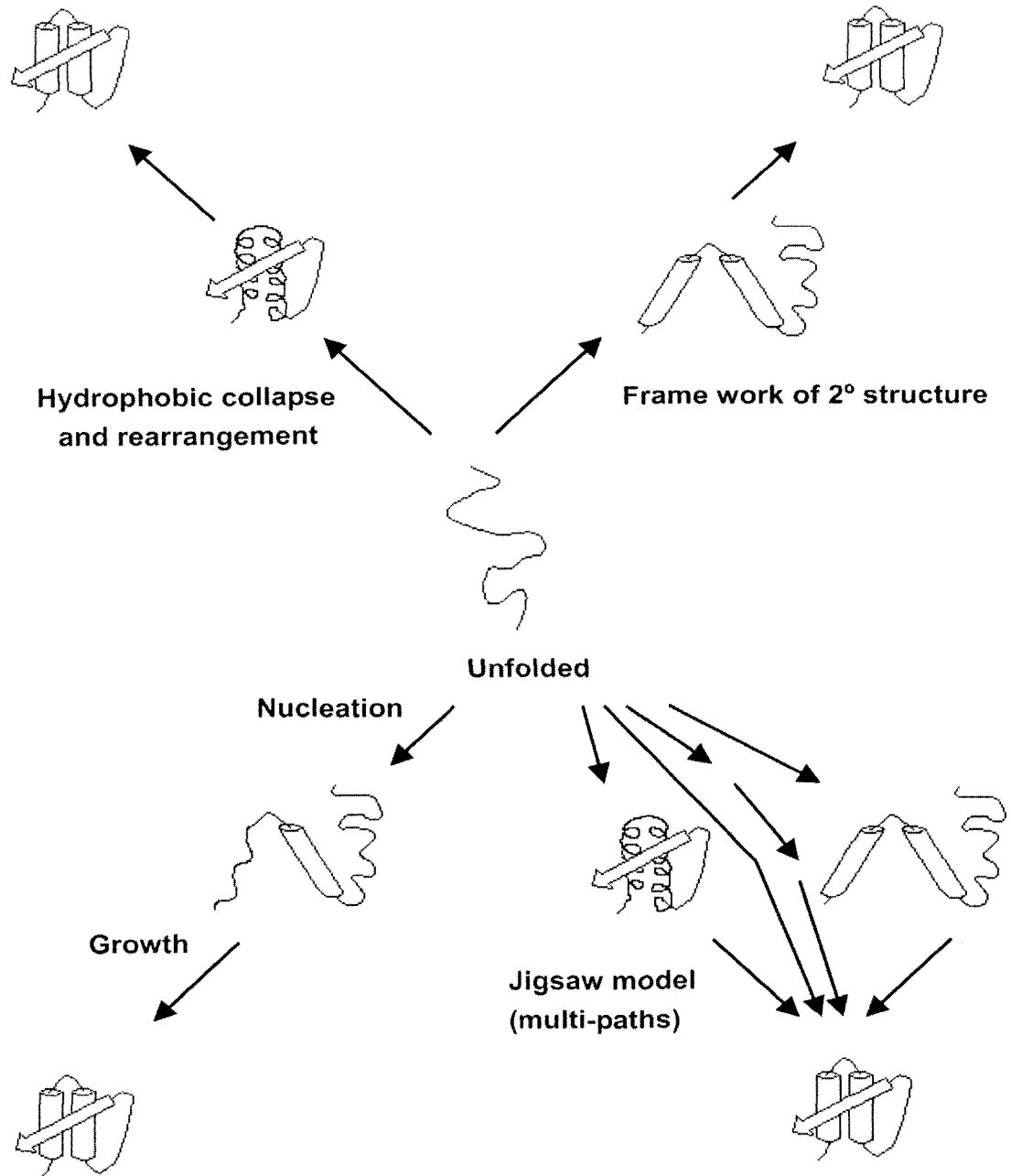


Figure 1. Classical models for mechanisms of protein folding.