이종 단백질 기능적 과생산을 위한 유전자 재설계: 동의 코돈(synonymous codon)을 활용한 번역속도조절

전남대학교 김근중 교수, 임호동 박사과정

1. 개요

대용량/고속 차세대 유전체 분석기법의 발달과 유전자 합성비용의 감소로 인해 의약이나 산업분야에 활용가치가 큰 고부가 단백질이나 대사물질의 발굴(mining)과 생산에 필요한 제반 여건이 마련되어지고 있다. 근본적으로, 대사물질도 단백질인 효소활성의 결과물임으로, 이들 자원개발의 선결요건은 고수율로 기능성 단백질을 대량으로 확보하는 일일 것이다. 하지만 재조합 단백질의 발현에는 수많은 변수가 상존하고 있기 때문에 예기치 못한 원인으로 인해 이종숙주에서의 재조합 단백질 과생산은 해결해야할 많은 난제들이 남아 있다. 공학적인 관점에서 살펴보면, 많은 변수에 의한 예측 불가능한 실패는, 생체내에서 단백질을 생산하고 관리(QC)하는데 이용하는 발현논리나 품질관리 방법, 그리고 다양한 생물리화학적 변수에 대응하는 생리기작들을 제대로 이해하지 못하고 있는 것이 가장 큰 이유일 것이다. 따라서 생체내의 발현논리나 생리기작에 근거해 숙주의 스트레스를 줄이며 발현품질을 보장받을 수 있는 새로운 재조합 단백질 대량생산 기술이 지속적으로 요구되고 있다. 이러한 목적에 부합하는 가장 효율적인 접근법은 발현논리나 품질제어, 생리기작을 고려한 발현유전자의 재설계일 것이다.

이종숙주에서 재조합 단백질의 생산을 증가시키기 위한 유전자 재설계의 주요 접근 방법 중 하나는, 코돈 최적화(codon optimization)와 같이 종간에 다른 빈도(선호도에 따른)로 존재하는 동의 코돈을 대체 도입하는 것이다. 단백질 번역속도의 증가에 주안점을 두고 있는 일방적인 최적화 (비선호코돈을 선호코돈으로 변경)와 달리, 최근에는 번역속도의 감소나 가속과 같은 새로운 발현논리에 근거한 유전자 재설계 기술들이 시도되고 있다. 유전자 내 동의코돈의 치환은 번역되는 아미노산 서열을 변경시키지 않고 암호화된 단백질의 생합성 속도와 효율을 조절할 수 있다. 본 글에서는 번역속도 조절을 위한 동의코돈 활용 유전자 재설계에 대한 일반적인 접근방법을 논의하고, 재조합 단백질 생산수율 향상과 기능적 발현의 다양한 측면에 대한 동의코돈 치환효과에 관련한 최근 연구동향을 기술하였다.

2. 동의코돈이 단백질 번역에 미치는 영향

분자생물학의 중심원리(central dogma)인 DNA에서 RNA로 전사되어 단백질로 번역되는 일련의 과정은, tRNA가 코돈(3개의 염기로 구성된 아미노산 지정 암호)에 대응하는 아미노산의 직접적인 연결을 제공하여 단백질생성에서 주된 역할을 담당한

다. 유전암호는 20개의 아미노산을 암호화하는 61개의 코돈과 3개의 번역 종결신호를 암호화하는 코돈으로 구성되며, 메티오닌(Met)과 트립토판(Trp)를 제외한 18개 아미노산은 여러 개의 동의코돈에 의해 암호화 된다 - 최근에 확인된 21번과 22번 아미노산인 selenocysteine 과 pyrrolysine 은 일반적(범용적)이지 않음으로 논의에서 배제하였다. 동의코돈을 구성하는 염기서열의 처음 두 개는 일반적으로 동일하지만 세 번째 염기서열은 선택성이 떨어져, 이를 처음 발견한 과학자들은 코돈에는 중복성(redundancy)이 있는 것으로 생각하였다. 또한 이를 단순히 코돈의 퇴화(degenerate)과정으로 생각하여 진화가 일어나는 하나의 기작으로 받아 들였다. 따라서 하나의 아미노산을 지정하는 여러 개의 코돈이 존재하는 현상을 진화와 종간의 특성을 구분하는 선호도(preference)나 tRNA 유전자나 aminoacyl tRNA synthetase 와의 동시진화 (coevolution) 산물로 생각하는 구조적 signature 로서의 특성으로 인식하였다. 하지만 이후 일련의 연구를 통해, 코돈은 번역시 필요한 아미노산을 단순지시하는 역할 이외에, 번역인자와의 상호작용은 물론 전사인자와의 상호작용(원핵은 전사와 번역이 coupling)에도 관여하며, 번역과정에서 단백질의 구조적 접힘이나 구획화, 그리고 분비에 관여하는 인자와의 상호작용에 필요한 번역속도 조절에도 참여하는 것으로 밝혀지고 있다(1, 2). 동의코돈의 치환은 암호화된 아미노산 서열 자체에는 영향을 주지 않기 때문에 발현 단백질 본연의 기능에는 영향을 미치지 않지만 이종간의 유전체 서열분석 비교에 따르면 각 종간 선호되는 코돈 빈도에 상당한 차이가 존재하기에 번역효율이나 품질관리, 전사체나 번역물의 안정성에 직접적으로 관여하는 것으로 추정된다(그림 1).

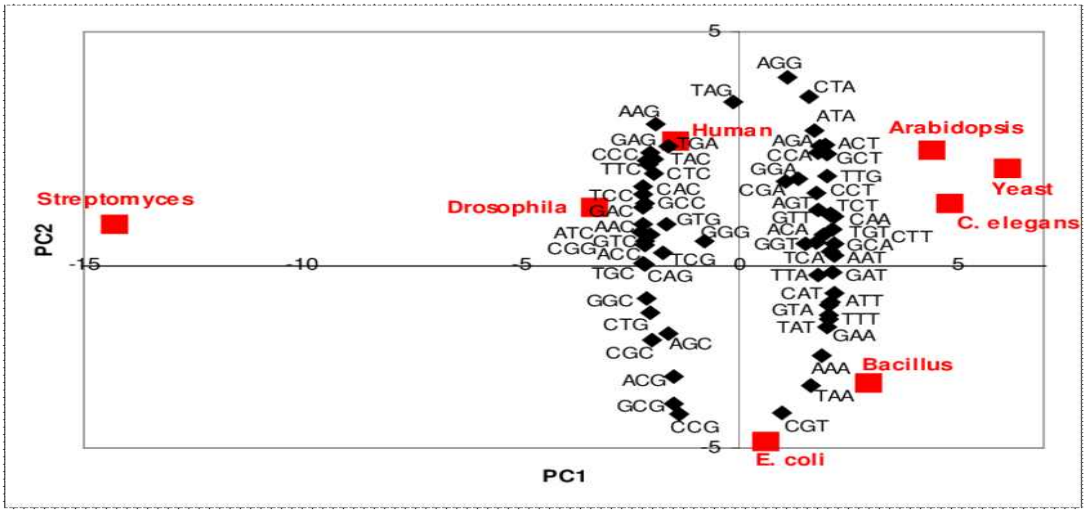

그림 1. 종간 코돈사용 빈도 비교. 원핵 미생물인 대장균을 비롯하여 인간 유전체까지 총 8종 게놈의 평균 코돈 선호도를 분석함. *Streptomyces* 는 가장 극단적인 코돈사용 빈도를 보여주며, *Yeast, C. elegans* 그리고 *Arabidopsis* 의 경우 유사한 양상을 보여줌. 이를 통해 전형적인 관점에서 보면 *Yeast* 가 *C. elegans* 나 *Arabidopsis* 유래의 이종 단백질을 생산하기에 좋은 숙주로 예상할 수 있음 [*Trends Biotechnol 22, 346-353 (2004)*]. 하지만 코돈빈도외에 순서(order)나 분포패턴은 비교되지 않아 좀 더 심도 있는 비교는 불가능함.

하지만, 제시된 결과에서는 번역과정에 대한 코돈의 (숨겨진)기능이나 효과를 유추할 수 있는 개연성 있는 기작에 대한 설명은 제시되어 있지 않다. 주된 주장으로는 각각의 코돈에서의 번역 신장율은 tRNA의 공급에 달려 있으며, 느려진 번역 신장율은 번역효율을 감소시킨다는 것이다. 따라서 일반적으로 번역율이나 수율감소는 희귀 tRNA에 의해 서서히 번역되는 개별코돈의 작은 효과의 누적에 의한 것이라고 제안하고 있다. 하지만 코돈사용의 단순한 빈도와 tRNA copy 수를 비교하는 이러한 단차원적인 해석은 경계해야할 여지가 많다. 게놈상의 희귀코돈(비선호 코돈)이라도 염색체상에 존재하는 tRNA 의 copy 수가 적지 않은 경우가 많으며, 주기적으로 반복되어 사용하는 희귀코돈은 번역속도를 저해하지 않는다는 연구결과도 존재하기 때문이다. 하지만 이러한 결과물과 관련된 정보들로 인해 유전자내의 비선호코돈을 선호코돈으로 대체하는 코돈 최적화 기술이 단백질 발현양을 증대시킬 것이라는 기대로 인해 관련된 많은 시도들이 지속적으로 보고되고 있다. 아래 예에서는 이러한 논리에 부분적으로 혹은 전적으로 대비되는 비선호코돈 사용과 분포의 다른 의미에 대해 논의한다.

2.1. 유전자 양 말단에서의 희귀코돈 분포와 의미

단백질 정보가 암호화 되어 있는 ORF에서의 동의코돈 빈도나 순서 등은, 무작위적으로 구성된 것이 아닌 코돈선택에 따른 진화론적 압력의 결과물로 판단된다(3-5). 대장균을 비롯한 많은 원핵생물 유전자를 비롯하여 효모와 같은 일부 진핵세포 유전자의 5'-말단에서 희귀코돈(rare codon)이 모여 있는 현상이 발견되었으며, 대략 20-30개 내외의 전형적인 말단의 코돈위치에서 발견되는 희귀코돈의 군집화는 생리나 유전자의 발현율에 근거한 의도적인 선택에 의한 영향임을 판단할 수 있는 근거를 제시하였다(6). 일반적으로 많은 미생물의 희귀코돈은 A/T 염기의 빈도가 높다고 알려져 있다. 따라서 5'-말단 위치의 희귀코돈이 모여 있는 형태는 mRNA의 해당 영역에서 2차 구조 형성이 감소되는 효과를 가져와 RBS의 용이한 노출로 인해 리보솜의 접근이 쉬워지게 된다. 그러나 5'-말단의 희귀코돈 군집화 형태는 전체 유전체의 GC 함량이 최소 50% 이상인 미생물에서 주로 발견되었다는 점에 유의해야 한다(7). 이와 더불어 매우 흥미롭게도, 분비 단백질을 코딩하는 유전자의 5'-말단 위치에서도 희귀코돈의 빈도가 높은 것이 관찰되었다. 이러한 특성이 번역 초기단계에서 일시적인 번역속도의 감소를 초래해 분비 단백질이 세포막으로 이동하는 메커니즘을 도와 외분비 효율을 증진 시킨다는 연구결과가 제시되었다. 말단영역에서의 번역속도 감소는 외분비나 구획화(compartmentalization)에 매우 중요한 샤페론과 같은 보조인자와의 상호작용 가능성을 높일 수 있기 때문이다. 예를 들어 효모의 경우, 신호서열 혹은 막횡단 영역근처의 ~35에서 40 코돈 위치에서의 희귀코돈 존재가 번역속도를 국지적으로 감소시켜 SRP(signal recognition particle)에 의한 초기 번역 펩타이드의 인식을 촉진시킨다고 한다(8). 이와 유사하게 대장균 분비단백질의 ORFeome을 분석해 보면, Shine-Dalgarno 서열 근처에서 일시적으로 번역속도를 감소시키는 서열이 존재하

여, 세포막으로의 구획화나 관통을 위한 요소 (factor)를 효율적으로 활용할 수 있는 기반을 제공할 수 있다고 보고되었다(9).

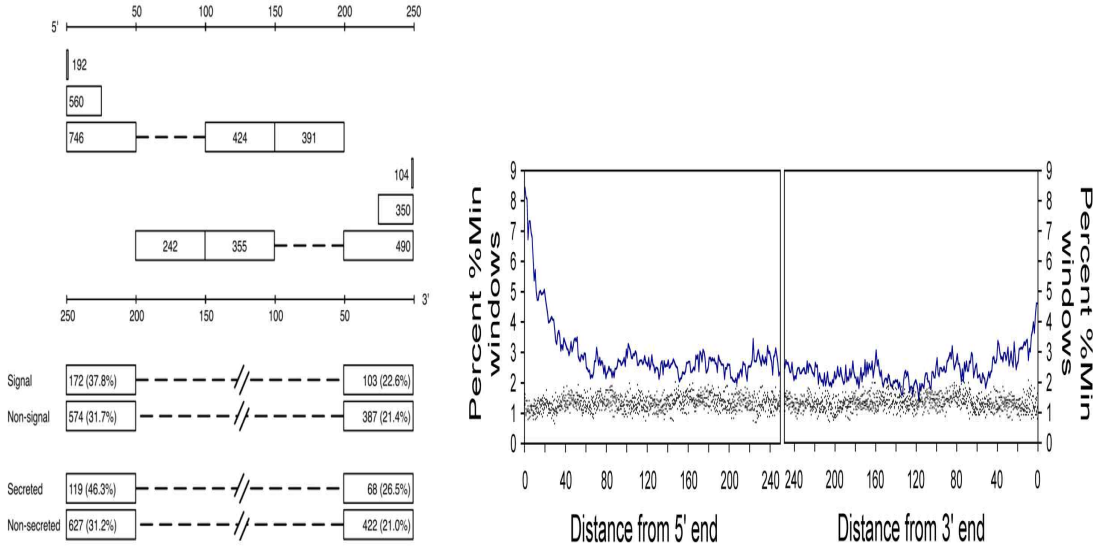

그림 2. 대장균 유전자중 양말단 부위에 집중적으로 분포하는 희귀코돈이 존재하는 유전자 비율(왼쪽)과 도식화된 특성(오른쪽). 출처: BMC Genomics 2010, 11:118.

대장균과 일부 원핵생물 ORF의 3'-말단에서도 이와 같은 희귀코돈 분포가 발견되었으나 mRNA의 분해를 막기 위한 기능(말단에서의 번역속도 감소는 naked mRNA가 가수분해효소에 노출되는 정도를 낮춤)이나 번역 종결단계를 확실하게 하기 위한 수단, 리보솜 분리 전까지 단백질 구조접힘의 용이성 제공이나 prematured polypeptide 와 상호작용하는 인자와의 접촉빈도 증가 등의 기능제공으로 해석되지만, 현재까지의 연구결과로는 5'-말단과 같은 뚜렷한 근거를 제시하기 어렵다. 상기 설명에서 볼 수 있는 것처럼 아직까지는 이론적인 설명에 국한되지만, 논리적으로 타당한 관점에서 mRNA의 5'-(ORF)말단에 희귀코돈(또는 AT 비율이 풍부한 코돈)을 전략적으로 배치함에 따라 유전자 재설계를 통한 단백질 생산수율 향상 및 분비율 증진, 단백질 구조접힘을 효율적으로 유도할 수 있을 가능성이 있는 것으로 기대된다.

2.2. ORF 내부에 존재하는 희귀코돈의 분포와 역할

일반적으로 알려진 것처럼 자주 사용되는 선호코돈은 희귀코돈보다 상대적으로 빠르게 번역될 가능성이 높다. 생물 종 마다 같은 아미노산을 지정하는 동의코돈간의 희귀성과 선호 정도는 매우 다르지만, 각 종마다의 희귀코돈 분포가 무작위적으로 분포하지 않고 전략적으로 배치되어 있는 것처럼 보이는 패턴이 ORF 내에서도 흔히 관찰된다. 이러한 희귀코돈이 국지적으로 모여 있는 형태는 발현율이 높고 낮음

(proteome 결과를 기준으로)에 상관없이 모든 유형의 단백질에서 폭넓게 확인되어진다. 단순한 유전체 염기서열이나 분석된 게놈정보와의 비교를 통해서도 희귀코돈의 분포나 배치되어 있는 형태는 특정 유전자에 국한되지 않으며 게놈 전체 GC 함량과도 무관하게 관찰된다(10, 11). 이러한 특이한 패턴은 유전자 발현에 있어 희귀코돈 분포의 중요성을 예측할 수 있는 중요한 근거를 제시한다. 일부 연구에서는 이러한 패턴이 번역과정에서 순차적으로 발생되는 구조적 접힘이 용이하도록 리보솜의 번역 속도를 감소시켜 적절한 단백질 구조형성을 유도하는 역할을 한다고 제안하였다(3, 12). 또한, 매우 희귀한 코돈이 상대적으로 단백질 도메인과 같은 큰 구조단위의 경계에서 확인되는 반면, 이와 상대적으로 덜 희귀한 코돈의 경우 단백질 모티프나 서브도메인과 같은 작은 단위의 경계에서 발견되는 것으로 보아 유전자에 존재하는 희귀코돈의 분포패턴이 단백질 접힘현상과 관련되어 모든 생물체가 공통적으로 이용하는 전략으로 활용될 가능성이 있어 보인다(13). 따라서 동의코돈을 활용한 희귀코돈 집단의 ORF내 전략적 배치는 단백질 생합성 과정에서 적절한 구조적 접힘을 유도하여 기능적으로 안정한 단백질 발현에 영향을 주는 것으로 판단되며, 이러한 가설은 이미 다국적 기업이나 미국의 일부 생명공학 회사에서 실제 응용하는 것으로 알려져 있다.

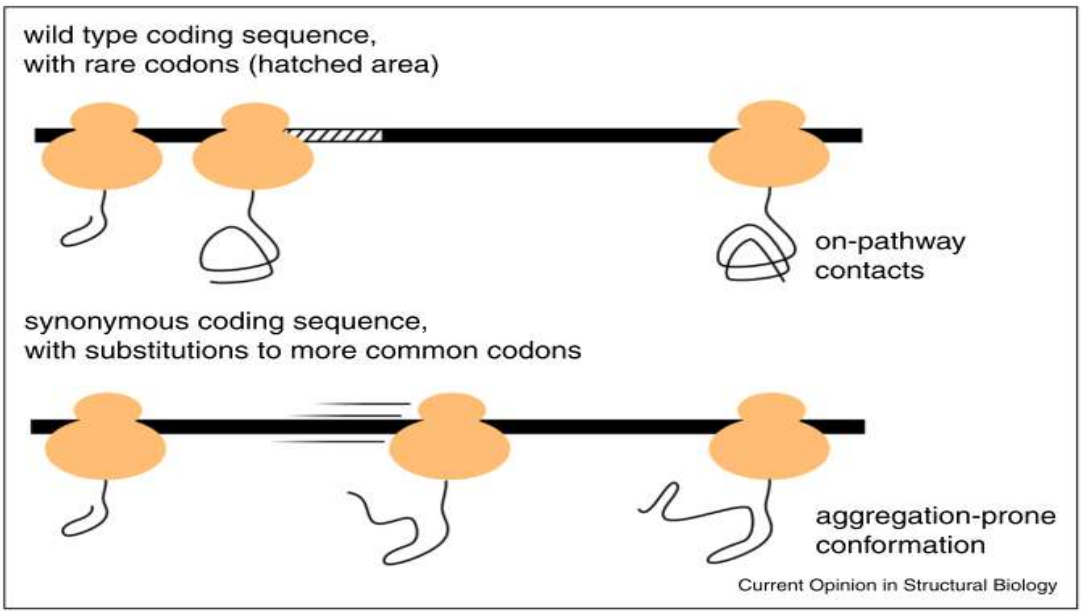

그림 3. ORF 내 선호코돈 빈도에 의한 번역속도 조절이 단백질 구조접힘에 주는 영향. 동의코돈을 이용하여 유전서열 내에 선호코돈 혹은 희귀코돈을 삽입 한 결과, 적절한 희귀코돈의 적용은 번역과정에서 리보솜의 속도가 조절되어 중요한 단백질 구조형성이 요구되는 구간에서 정확한 단백질 구조 접힘을 유도할 수 있으나 전체 유전서열을 선호코돈으로 치환 할 경우, 번역 속도는 향상되지만 단백질 구조형성에 요구되는 시간이 부족하여 불용성으로 발현될 가능성이 높음

3. 동의 코돈과 mRNA 안정성

mRNA의 안정성은 유전자 발현조절에 중요한 역할을 한다. 단순하게 생각하면 세포내에서 오래 유지되는 mRNA가 그렇지 못한 mRNA 보다 일반적으로 더 많은 단백질 생성에 주형으로 작용할 수 있는 개연성이 크다. 코돈분포나 선호도에 의한 mRNA 안정성의 변화에는 근래까지 뚜렷한 상관관계가 보고되고 있지 않았지만, 2015년 Cell에 투고된 논문에 의하면 mRNA의 열역학적 안정성보다 코돈 분포나 패턴에 기반한 (생물학적)최적화가 세포 내 mRNA 분해율에 광범위하게 영향을 미치는 것으로 제안되었다(14). 이들은 효모를 이용한 연구를 통하여 동의코돈 치환에 의한 원형 유전자(original)의 코돈최적화가 mRNA의 안정성을 높여 세포질에서 오랜 시간 유지되었음을 확인하였다. 또한 흥미롭게도 이러한 결과가 대장균에서도 확인되었다. 따라서 리보솜에 의한 mRNA의 번역속도를 변화시키면 더 안정적으로 많은 단백질을 생산할 수 있음을 제안하였다(15). 상기 논문에서와 같이 번역속도의 증대가 전사체의 안정성이나 이에 따른 단백질 생성량을 일반적으로 증가시킬 수 있을 것이라는 관점에는 일부 논란의 여지가 남아있지만, 빠르건 느리건 간에 속도조절이 지닌 잠재적 가능성을 예측하기에는 큰 무리가 없어 보인다. 결국, 어떠한 논리가 더 지배적인 확장성을 지니는지를 확인하는 작업과 이러한 논리가 지닌 범용성을 확인하는 과정을 거쳐 희귀코돈과 선호코돈의 적절하게 조합하여 유전자를 재설계하면 목적 단백질의 생산성 향상을 유도할 수 있을 것이라는 결론에 도달할 수 있을 것이다.

4. 고찰

동의 코돈을 이용한 유전자 재설계는 유용 단백질의 기능적 과생산을 위한 유전서열의 최적화 방안의 하나이다. 여기서 최적화란 단순히 발현속도를 높이려는 희귀코돈의 배제를 의미하지 않는다. 진화적으로 채용한 세포내 발현논리나 전략을 모티프로 이용하는 공학적 모방을 의미한다. 이러한 접근법은 이종 숙주에서 외래 유전자를 발현시키는 과정에서 난발현성을 보이거나 불용성으로 발현되는 경우, 이들의 발현율을 증진하고 단백질 구조 접힘이 용이하도록 재설계가 가능하다는 것을 의미한다. 동의코돈 기반 유전자 재설계에 의한 단백질생산 최적화는 산업용 효소는 물론, 아미노산 서열변화는 유발하지 않는 유전서열 조작법이기 때문에 인체에 적용되는 의약 목적의 치료용 단백질을 고효율로 생산할 수 있는 잠재성을 지니고 있다. 또한 이러한 접근법을 통해 구축된 동의코돈 대체 라이브러리의 분석을 통해 단백질 발현이나 품질유지, 기능조절에 관련된 코돈의 영향을 이해할 수 있을 것이다.

<참고문헌>

1. A. B. Stergachis *et al.*, Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367-1372 (2013).

2. D. J. D'Onofrio, D. L. Abel, Redundancy of the genetic code enables translational pausing. *Front Genet* **5**, 140 (2014).
3. J. L. Chaney, P. L. Clark, Roles for Synonymous Codon Usage in Protein Biogenesis. *Annu Rev Biophys* **44**, 143-166 (2015).
4. J. B. Plotkin, G. Kudla, Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12**, 32-42 (2011).
5. J. L. Chaney, P. L. Clark, Roles for Synonymous Codon Usage in Protein Biogenesis. *Annu Rev Biophys* **44**, 143-166 (2015).
6. D. B. Goodman, G. M. Church, S. Kosuri, Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475-479 (2013).
7. M. Allert, J. C. Cox, H. W. Hellinga, Multifactorial Determinants of Protein Expression in Prokaryotic Open Reading Frames. *Journal of Molecular Biology* **402**, 905-918 (2010).
8. S. Pechmann, J. W. Chartron, J. Frydman, Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat Struct Mol Biol* **21**, 1100-1105 (2014).
9. N. Fluman, S. Navon, E. Bibi, Y. Pilpel, mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. *Elife* **3**, (2014).
10. T. F. Clarke, P. L. Clark, Rare Codons Cluster. *Plos One* **3**, (2008).
11. M. Chartier, F. Gaudreault, R. Najmanovich, Large-scale analysis of conserved rare codon clusters suggests an involvement in co-translational molecular recognition events. *Bioinformatics* **28**, 1438-1445 (2012).
12. E. P. O'Brien, P. Ciryam, M. Vendruscolo, C. M. Dobson, Understanding the Influence of Codon Translation Rates on Cotranslational Protein Folding. *Accounts Chem Res* **47**, 1536-1544 (2014).
13. A. A. Komar, A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* **34**, 16-24 (2009).
14. V. Presnyak *et al.*, Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111-1124 (2015).
15. G. Boel *et al.*, Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature* **529**, 358-+ (2016).