Kód předmětu: C8980

MASARYKOVA UNIVERZITA

Protein expression and purification

- IV. DNA cloning

Lubomír Janda, Jozef Hritz, Blanka Pekárová, Radka Dopitová and Adam Norek

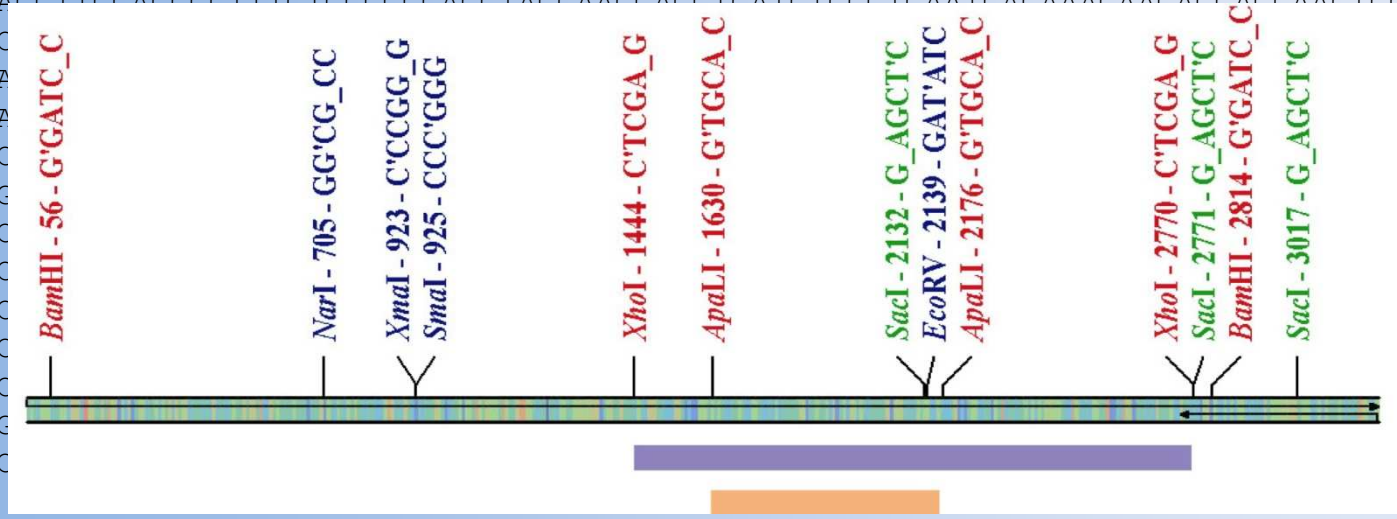
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

EVROPSKÁ UNIE esf MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

4.1. Introduction: correctness of your construct – cloning strategy

ATGGGCGGCATccACAGGGTGAACAGATGTACCGGAGGGTGTATCGTCTGCATGAGCGCCTGGTAGCCATCCGCACTGAGTACAACCTCC
GGCTGAAGGCAGGAGTGGGTGCCCCTGTGACCCAGGTGACCCTGCAGAGTACACAGAGGCGCCCAGAGCTAGAGGACTCCACACTGCGCT
ACCTGCAAGACCTGCTGGCCTGGGTAGAGGAGAACCAGCGTCGAATAGACAGTGCTGAGTGGGGCGTGGACTTGCCCAGTGTGGAGGCCC
AGCTGGGCAGCCACCGAGGCATGCATCAGTCTATAGAGGAATTTCGGGCCAAGATCGAGCGGGCTCGGAATGATGAGAGCCAGCTCTCCC
CTGCCACCCGGGGTGCCTACCGGGACTGCCTAGGTCGCCTAGACCTGCAGTATGCAAAGCTGCTGAACTCCTCCAAGGCCCGCCTCCGGT
CCCTGGAGA
GGAGTGACC
AGATCCAGA
GGAGCTGGA
CGGAGGAAC
TGCAGGATG
TGAAGCCAC
AGGGTGACC
TGTGCTTTC
ACCAGCTTC
TCCGCACAC
GTGGCTTTG
AGGGTGAGC



AAAATCAAGG
CAGACACAGT
GTTCGGGAGG
GAGgATCTGC
ATTGTGCAGT
AcTGTGCACA
GTGCCTTCTG
ACACTGTGGC
CTAGTCACGT
CAGGATGCCG
AGCCTGGAGC

DGVRANELQLRWQEYRELVLLLLQWIRHHTAAFEERKFPSSFEEIEILWCQFLKFKETELPAKEADKNRSKVIYQSLEGAVQAGQLK
IPPGYHPLDVEKEWGKLHVAILEREKQLRSEFERLECLQRIVSKLQMEAGLCEEQLNQADALLQSDIRLLASGKVAQRAGEVERDLD
KADGMIRLLFNDVQTLKDGRHPQGEQMYRRVYRLHERLVAIRTEYNLRLKAGVGAPVTQVTLQSTQRRPELEDSTLRYLQDLLAWVE
ENQRRIDSAEWGVDLPSVEAQLGSHRGMHQSIEEFRAKIERARNDESQLSPATRGAYRDCLGRLDLQYAKLLNSSKARLRSLESLHG
LQLCCCIEAHLKENTAYFQFFSDVREAEEQLQKLQETLRRKYSCDRTITVTRLEDLLQDAQDEKEQLNEYKGHLSGLAKRAKAIVQL
VEECQKFAKQYINAIKDYELQLITYKAQLEPVASPAKKPKVQSGSESVIQEYVDLRTRYSELTTLTSQYIKFISETLRRMEEEE

Key concepts:
- Knowing the objectives before DNA cloning
- Appreciating the complexity of plasmid systems in terms of the numbers of distinct plasmid options

4.2. The key questions before DNA cloning

4.2.1. DNA-protein analysis

4.2.1.1. Plasmid map

DNA sequence

```
ATGGCTAGCACAGATTCAGAGAGTGAGACTAGGGTCAAGTCAGTGCGTACCGGTCGAAAG
CCTATTGGGAACCCAGAGGACGAGCAAGAGACTTCCAAGCCGAGTGACGATGAATTCTTA
AGAGGAAAGAGAGTTCTTGTGGTCGATGATAACTTTATATCACGTAAAGTTGCAACAGGA
AAGCTGAAGAAGATGGGAGTCTCAGAGGTCGAACAATGCGACAGTGGGAAAGAAGCTTTG
AGATTAGTCACTGAAGGGCTTACACAAAGAGAAGAACAAGGTTCAGTAGATAAACTTCCG
TTTGACTACATATTCATGGACTGCCAAATGCCAGAAATGGATGGCTATGAAGCAACTAGA
GAGATTAGGAAAGTGGAGAAAAGTTATGGGGTGCGTACACCAATTATAGCTGTATCTGGT
CATGATCCTGGTTCAGAGGAAGCAAGAGAAACCATTCAAGCTGGAATGGACGCCTTCTTA
GATAAAAGCTTGAATCAACTTGCAAACGTCATTAGAGAAATCGAAAGCAAACGTCAC
```

www.expasy.ch translate

```
MASTDSESETRVKSVRTGRKPIGNPEDEQETSKPSDDEFLRGKRVLVVDDNFISRKVATG
KLKKMGVSEVEQCDSGKEALRLVTEGLTQREEQGSVDKLPFDYIFMDCQMPEMDGYEATR
EIRKVEKSYGVRTPIIAVSGHDPGSEEARETIQAGMDAFLDKSLNQLANVIREIESKRH
```

4.2. The key questions before DNA cloning

4.2.1. DNA-protein analysis

4.2.1.2. Secondary structure prediction

www.expasy.ch jpred3

```
MASTDSESETRVKSVRTGRKPIGNPEDEQETSKPSDDEFLRGKRVLVVDDNFISRKVATG
--EEEEEE----EEEEEEEEEE--------------------EEEEEE--HHHHHHHHH

KLKKMGVSEVEQCDSGKEALRLVTEGLTQREEQGSVDKLPFDYIFMDCQMPEMDGYEATR
HHHH----EEEEE--HHHHHHHHHH----------------EEEEEE------HHHHH

EIRKVEKSYGVRTPIIAVSGHDPGSEEARETIQAGMDAFLDKSLNQLANVIREIESKRH
HHHH---------EEEEEE-----HHHHHHHHHH-----E---HHHHHHHHHHHHH---
```

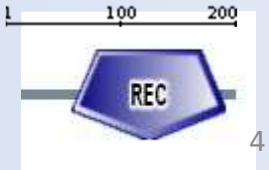
4.2.1.3. Domains detected by SMART

www.expasy.ch SMART

```
KRVLVVDDNFISRKVATGKLKKMGVSEVEQCDSGKEALRLVTEGLTQREEQGSVDKLP
FDYIFMDCQMPEMDGYEATREIRKVEKSYGVRTPIIAVSGHDPGSEEARETIQAGMDA
FLDKSLNQLANVI
```

Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
REC	43	171	$1.19e^{-26}$

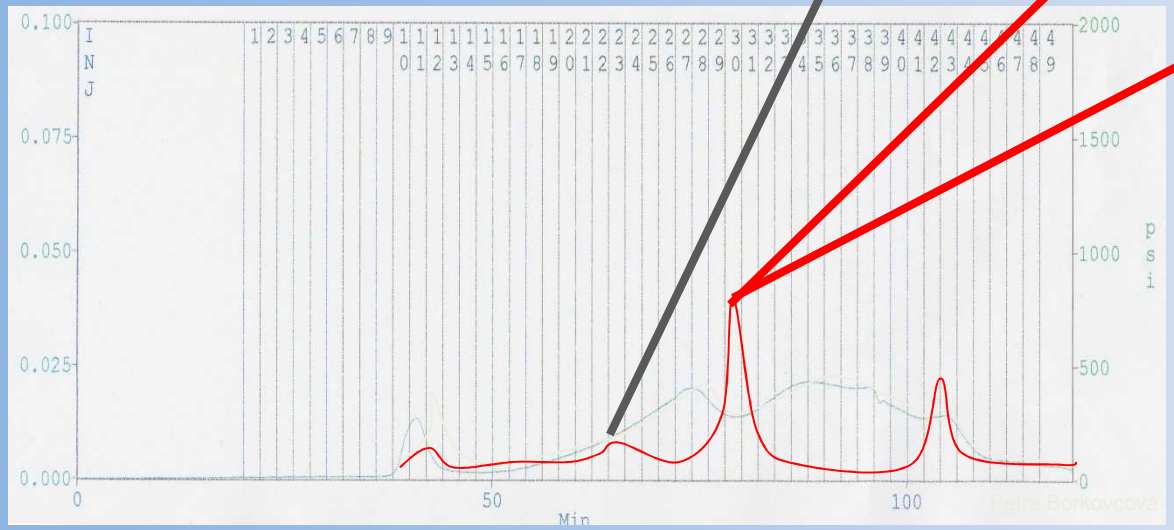


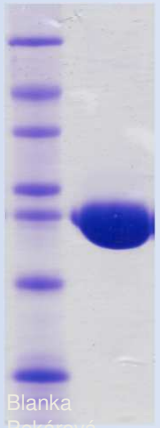
4.2.1.3. Domains detected by SMART

Histidine kinase from *A. thaliana* - CKI 1

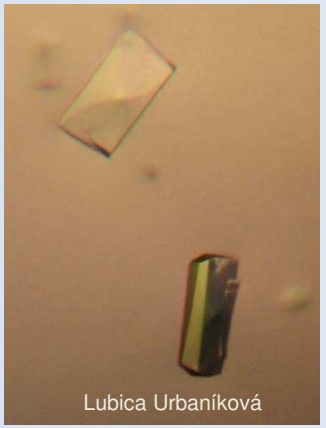
Three constructs of receiver domain.

pET 28





Blanka Pekárová



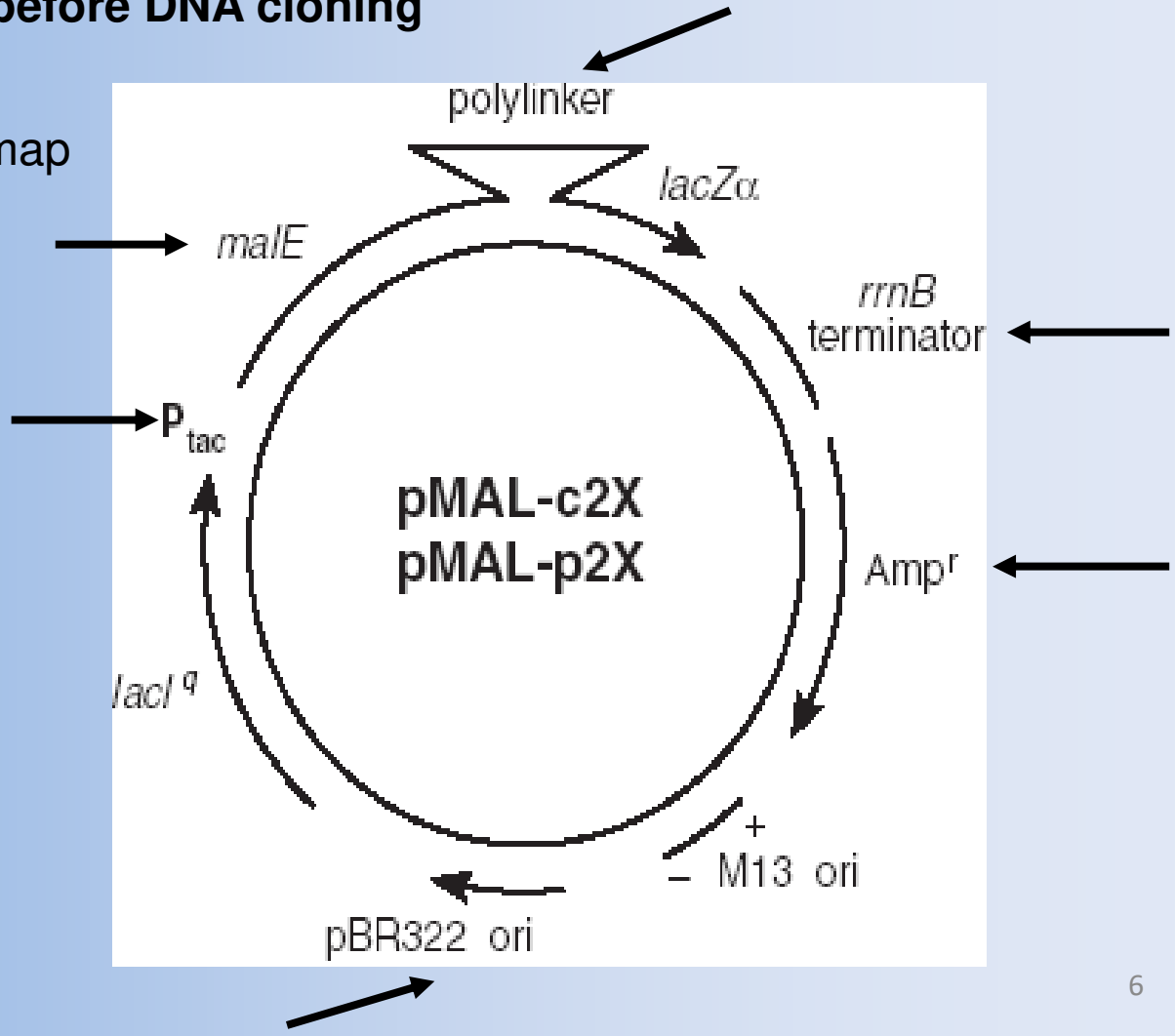
Lubica Urbaníková

Key concepts:
- Knowing the objectives before DNA cloning
- Appreciating the complexity of plasmid systems in terms of the numbers of distinct plasmid options

4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.1. Plasmid map



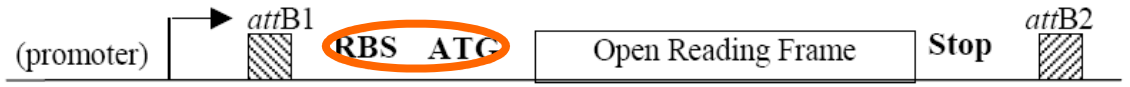
4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.1. Plasmid map

- Strong promoter — ptac, ptrp, λpL, pT_7
- Promoter regulation — ptrp-tryptophan/IAA; ptac-IPTG; λpL – temperature; pT_7 – IPTG
- Transcription terminator — T_7 term, rrnT1,T2
- Ribosome binding site — AAGG (upstream of the AUG initiation)
- SD-AUG spacing and base composition — Spacing is crucial to high level expression. (optimal distance 6–10 bp, AT rich base composition)

A. Expression clone structure:



B. Expression clone sequence:

Shine-Dalgarno Kozak Open reading frame (amino end)

5' - ACA AGT TTG TAC **AAA AAA** GCA GGC TTC GAA GGA GAT AGA ACC **ATG** NNN NNN NNN ---

3' - TGT TCA AAC ATG **TTT TTT** CGT CCG AAG GTT CCT CTA TCT TGG TAC NNN NNN NNN ---

attB1 Translation start*

Open reading frame (carboxy end)

--- NNN NNN NNN **TAG** GAC CCA GCT TTC TTG **TAC AAA** GTG GT - 3'

--- NNN NNN NNN ATC CTG GGT CGA AAG AAC **ATG TTT** CAC CA - 5'

Translation stop *attB2*

4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.2. Promoters

- *lacUV5*, *tac* and *trc* promoters are repressed by the *lac* repressor (*lacI* or *lacI*^{q}) and induced with IPTG.
- *Trp* promoter is repressed by the *trp* repressor and induced with tryptophan (or indole-3-acetic acrylic acid).
- *T7* promoter requires expression of phage RNA polymerase (host strain usually contains this polymerase expressed from *lac UV5* promoter induced by addition of IPTG).
- P_L lambda phage promoter exhibits maximum expression when induced and has low basal expression when the *cI* repressor is present.

4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.2. Promoters

4.2.2.2.1. T7*lac* promoter

Relative basal uninduced expression levels of cloned β-galactosidase with various vector/host combinations

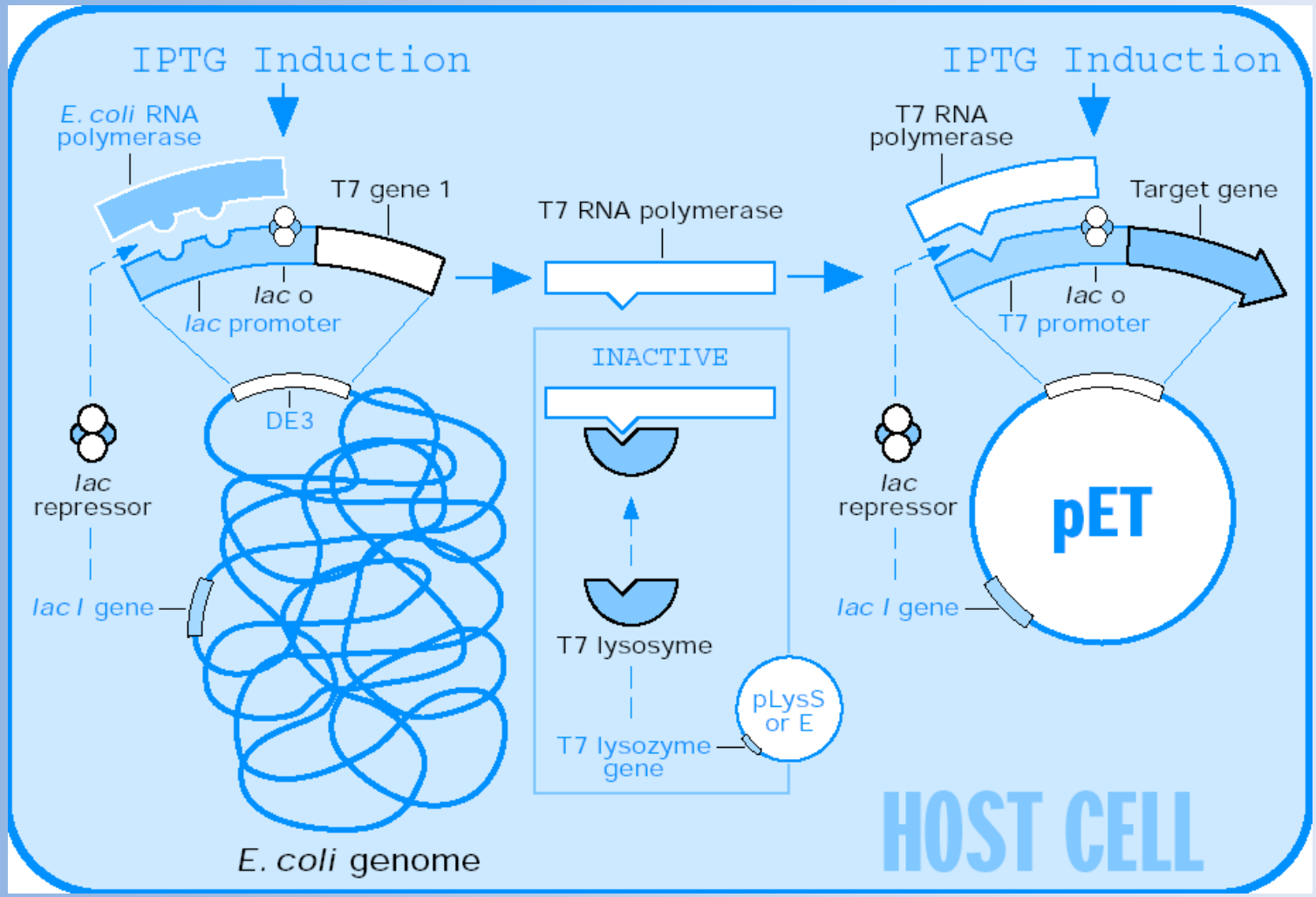
Promoter	T7	T7	T7	T7*lac*	T7*lac*	T7*lac*
Host	(DE3)	(DE3) pLysS	(DE3) pLysE	(DE3)	(DE3) pLysS	(DE3) pLysE
Activity	100%	30%	10%	10%	3%	1%

4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.2. Promoters

4.2.2.2.1. T7*lac* promoter



4.2. The key questions before DNA cloning

4.2.2. Plasmid option

4.2.2.3. Examples of *E. coli* expression systems

	Vector system	Promoter/ induction method	Special host protein tag strains required:	Source	Web site
	Pinpoint	*tac*/IPTG or *T7* IPTG	Yes	Biotin binding domain	www.promega.com
*	pET	*T7* IPTG	Yes	His_6, T7 gene 10	www.novagen.com
*	pGEX	*tac*/IPTG	No	GST	www.amershambiosciences.com
	pBAD	*ara*BAD	Yes	His_6, GFP	www.invitrogen.com
	pLEX	P_L/trp	Yes		
	pPROTet	P_{Ltet}/anhydrotetracyclin	No	His_6	www.clontech.com
	pTYB	*T7* IPTG	Yes	Chitin binding domain	www.neb.com
*	pMAL	*tac*/IPTG	Yes	Maltose binding domain	
*	pQE	*T5*/IPTG	Yes/TOPP	His_6	www.qiagen.com
	pCAL	*T7*/IPTG	Yes	Calmodulin binding peptide	www.stratagene.com
	pLAG	*tac*/IPTG	Yes		www.sigmaaldrich.com
	pUbEx15	tac/IPTG	Yes	ubiquitin	janda@vri.cz

4.2. The key questions before DNA cloning

4.2.3. N-terminal amino acids

N-terminal amino acids that reduce stability of proteins.

N-degrons

- Phe, Leu, Trp, Tyr and Arg, Lys,

Tobias et al, 1991, Science; Humbard et al., 2013, JBC

Amino acids stabilized in penultimate position
N-terminal methionin.

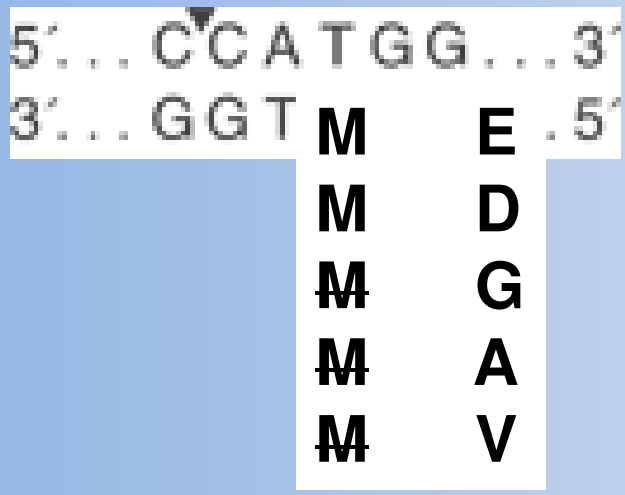
His, Gln,Glu, Phe, Met, Lys, Tyr, Trp, Arg

Hirel et. al., 1989, PNAS; Lathrop et al. 1992; Liao et.al., 2004, Protein Science

Methionione aminopeptidase remove the initiator Met in proteins when the second residue is

Gly, Ala, Ser, Cys, Thr, Pro or Val

Bonissone et al., 2013, Molecular and Cellular Proteomics



NcoI

NdeI

4.2. The key questions before DNA cloning

4.2.4. Protease recognition sites

Check the sequence of the fusion partner for the presence of additional protease recognition sites.

- Thrombin
  pH 8.0
  Pro-Arg/Gly
  Pro-Lys/Leu
  Ala-Arg/Gly
  Gly-Lys/Ala
  Ile-Arg/Ser
  Leu-Arg/Ala
  Ile-Arg/Ile
- PreScission
  pH 8.9
  Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro
- Factor Xa
  pH 6.5–7.5
  Ile-Glu-Gly-Arg/X
- Enterokinase
  pH 7.0–8.0
  Asp-Asp-Asp-Asp-Lys/X
- TEV protease
  pH 5.5 – 8.5
  Glu-Asn-Leu-Tyr-Phe-Gln/Ser

4.2. The key questions before DNA cloning

4.2.5. Antibiotic selection

- *bla* gene — ampicillin resistance

  Ampicillin x Carbenicilin

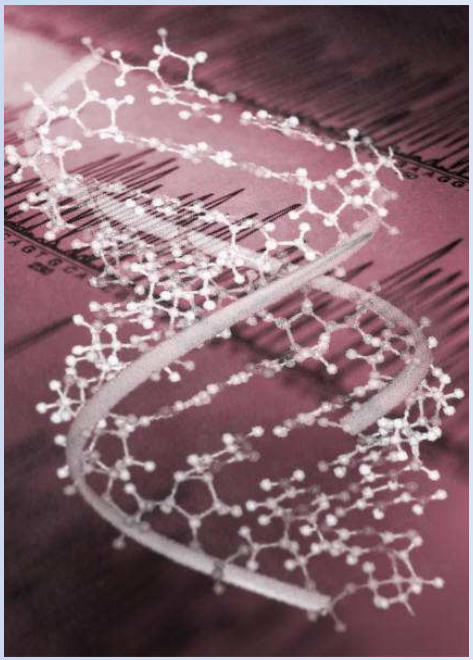
- *kan* gene — kanamycin resistance

4.2. The key questions before DNA cloning

4.2.6. Codons with translation problems

- Arginine AGG
  AGA
  CGA
  CGG
- Isoleucine AUA
- Leucine CUA
- Glycine GGA
- Proline CCC

BL21-Codon plus-RIL



BL21-Codon plus-RP

4.2. The key questions before DNA cloning

4.2.6. Codons with translation problems

http://www.kazusa.or.jp/codon/

Escherichia coli K12

UUU 19.7	UCU 5.7	UAU 16.8	UGU 5.9
UUC 15.0	UCC 5.5	UAC 14.6	UGC 8.0
UUA 15.2	UCA 7.8	**UAA** stop	**UGA** stop
UUG 11.9	UCG 8.0	**UAG** stop	UGG 10.7
CUU 11.9	CCU 8.4	CAU 15.8	CGU 21.1
CUC 10.5	CCC 6.4	CAC 13.1	CGC 26.0
CUA 5.3	CCA 6.6	CAA 12.1	CGA 4.3
CUG 46.9	CCG 26.7	CAG 27.7	CGG 4.1
AUU 30.5	ACU 8.0	AAU 21.9	AGU 7.2
AUC 18.2	ACC 22.8	AAC 24.4	AGC 16.6
AUA 3.7	ACA 6.4	AAA 33.2	AGA 1.4
AUG 24.8	ACG 11.5	AAG 12.1	AGG 1.6
GUU 16.8	GCU 10.7	GAU 37.9	GGU 21.3
GUC 11.7	GCC 31.6	GAC 20.5	GGC 33.4
GUA 11.5	GCA 21.1	GAA 43.7	GGA 9.2
GUG 26.4	GCG 38.5	GAG 18.4	GGG 8.6

Arabidopsis thaliana

UUU 21.8	UCU 25.2	UAU 14.6	UGU 10.5
UUC 20.7	UCC 11.2	UAC 13.7	UGC 7.2
UUA 12.7	UCA 18.3	**UAA** stop	**UGA** stop
UUG 20.9	UCG 9.3	**UAG** stop	UGG 12.5
CUU 24.1	CCU 18.7	CAU 13.8	CGU 9.0
CUC 16.1	CCC 5.3	CAC 8.7	CGC 3.8
CUA 9.9	CCA 16.1	CAA 19.4	CGA 6.3
CUG 9.8	CCG 8.6	CAG 15.2	CGG 4.9
AUU 21.5	ACU 17.5	AAU 22.3	AGU 14.0
AUC 18.5	ACC 10.3	AAC 20.9	AGC 11.3
AUA 12.6	ACA 15.7	AAA 30.8	AGA 19.0
AUG 24.5	ACG 7.7	AAG 32.7	AGG 11.0
GUU 27.2	GCU 28.3	GAU 36.6	GGU 22.2
GUC 12.8	GCC 10.3	GAC 17.2	GGC 9.2
GUA 9.9	GCA 17.5	GAA 34.3	GGA 24.2!
GUG 17.4	GCG 9.0	GAG 32.2	GGG 10.2

Leu-CUA 5.3/7.2/9.9
Ile-AUA 3.7/7.5/12.6
Pro-CCC 6.4/19.8/5.3 →
Gly-GGA 9.2/16.8/24.2 →

Arg-CGA 4.3/6.2/6.3
Arg-CGG 4.1/11.4/4.9
Arg-AGA → 1.4/12.2/19.0
Arg-AGG → 1.6/12.0/11.0

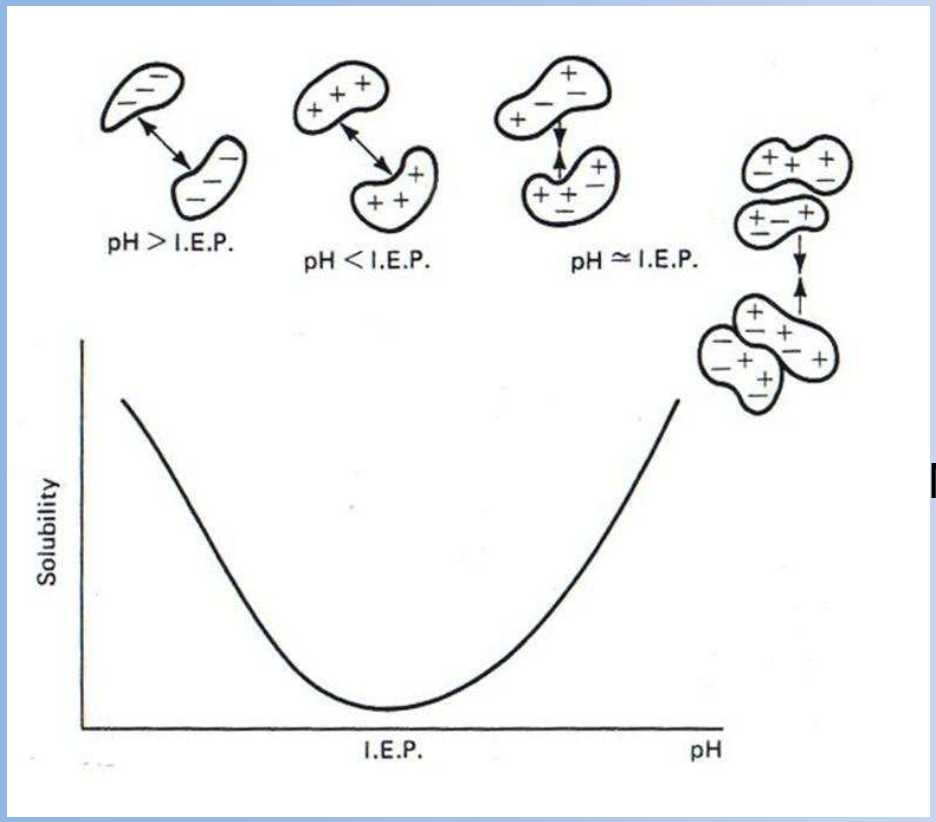
Key concepts: Being aware of solubility as a function of protein structure

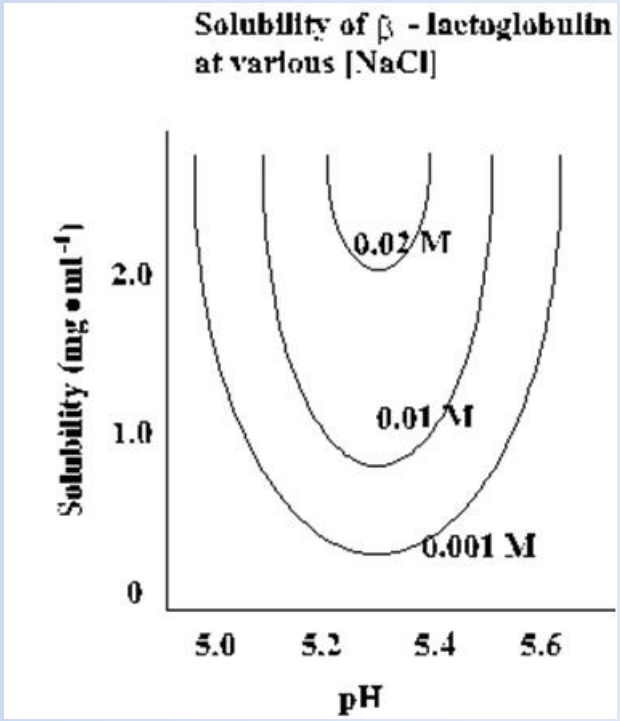
4.3. Protein solubility http://www.biotech.ou.edu/

- Low solubility in aqueous solvents is often regarded as an indication that a protein is “hydrophobic”.
- As native, properly folded structures aggregate less than unfolded, denatured ones, there is a close relationship between solubility and stability.
- The free energy of protein stabilization in an aqueous solution is very low (12 kcal/mol at 30°C).
- Free energy of unfolding is observed to be only 5–20 kcal/mol.
- Consequently, proteins are on the verge of denaturation.

4.3. Protein solubility

4.3.1. Determining surface charge





Most precipitation

$$PO_4^{3-} > SO_4^{2-} > COO^- > Cl^-$$

$$NH_4^+ > K^+ > Na^+$$

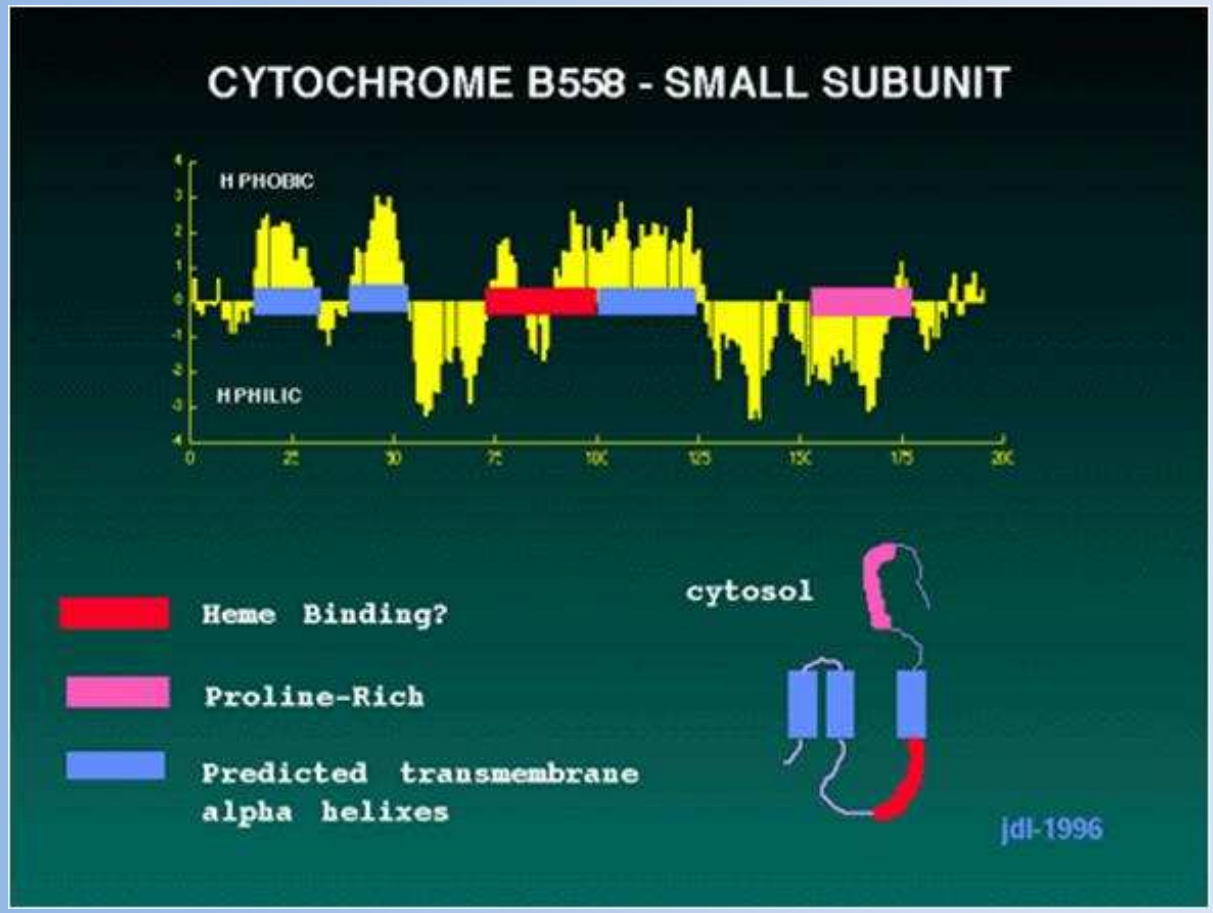
Least precipitation

Isoelectric focusing gives the pI, the pH at which the protein shows no net charge in isoionic conditions.
Generally, charged proteins can be “salted in” by counterions.

4.3. Protein solubility

4.3.2. Determining hydrophobicity

http://www.roselab.jhu.edu/~raj/MISC/hphobh.html



4.3. Protein solubility

4.3.3. Solubility model

http://www.biotech.ou.edu/

The revised Wilkinson-Harrison solubility model

$$CV = \lambda_1\left(\frac{N+G+P+S}{n}\right) + \lambda_2 \left|\left(\frac{(R+K)-(D+E)}{n} - 0.03\right)\right|$$

n	number of amino acids in the protein
N, G, P, S	number of Asn, Gly, Pro, or Ser residues
R, K, D, E	number of Arg, Lys, Asp, or Glu residues
λ_1, λ_2	coefficients (15.43 and -29.56)

The probability of the protein being soluble is based on the parameter CV – CV', where CV' is the discriminant, equal to 1.71.

If CV – CV' is **positive**, the protein is predicted **to be insoluble**, while if CV – CV' is **negative**, the protein is predicted **to be soluble**.
The probability of solubility or insolubility can be predicted from the following equation:

Probability of solubility or insolubility =

$$0.4934 + 0.276\,|(CV-CV')| - 0.0392\,(CV-CV')^2$$

4.3. Protein solubility

4.3.4. Protein engineering to increase solubility

4.3.4.1. Amino acid solubility and water affinity

- Hydrophobic amino acids cluster to avoid water.
- Most positively charged and amide side chain residues (His, Lys, Arg, Gln, Asn) were on the surfaces of the proteins studied.
- The interiors were primarily composed of aliphatics (Gly, Ala, Ile, Leu, Val, Phe).
- But only 23% of Trp residues and 13% of the Tyr in the structures were not accessible to the solvent, similar to that of the negative polar residues Glu (20%) and Asp (14.5%).

Amino acid		Transfer free energy kJ/mol	% buried
Phe	F	15.5	48%
Met	M	14.2	50%
Ile	I	13	65%
Leu	L	11.7	41%
Val	V	10.9	56%
Cys	C	8.4	47%
Trp	W	7.9	23%
Ala	A	6.7	38%
Thr	T	5	25%
Gly	G	4.2	37%
Ser	S	2.5	24%
Pro	P	-0.8	24%
Tyr	Y	-2.9	13%
His	H	-12.5	19%
Gln	Q	-17.1	6%
Asn	N	-20.1	10%
Glu	E	-34.3	20%
Lys	K	-36.8	4%
Asp	D	-38.5	15%
Arg	R	-51.4	0%

4.3. Protein solubility

4.3.4. Protein engineering to increase solubility

4.3.4.2. Peptide solubility

- For peptides of more than 8 amino acids, sequences favouring α-helix or random coil structures are more soluble in polar solvents than those forming β–sheet structures.

- For other peptides, insertion of arg-NO_2 residues, or replacement of hydrophobic residues, improved solubility and lowered aggregation tendencies.

Amino acid		Transfer free energy kJ/mol	% buried	Chou-Fasman coil index
Phe	F	15.5	48%	0.71
Met	M	14.2	50%	0.58
Ile	I	13	65%	0.66
Leu	L	11.7	41%	0.68
Val	V	10.9	56%	0.62
Cys	C	8.4	47%	1.18
Trp	W	7.9	23%	0.75
Ala	A	6.7	38%	0.7
Thr	T	5	25%	1.07
Gly	G	4.2	37%	1.5
Ser	S	2.5	24%	1.82
Pro	P	-0.8	24%	1.59
Tyr	Y	-2.9	13%	1.06
His	H	-12.5	19%	1.06
Gln	Q	-17.1	6%	0.86
Asn	N	-20.1	10%	1.35
Glu	E	-34.3	20%	1.2
Lys	K	-36.8	4%	0.98
Asp	D	-38.5	15%	1.2
Arg	R	-51.4	0%	1.04

4.3. Protein solubility

4.3.4. Protein engineering to increase solubility

4.3.4.3. Primary structure alterations

• **Replacement** of the hydrophobic
EGN**FF**GKIIDYIKLMFHHWFG
C-terminal amino acids of penicillin-binding protein 5 with a shorter hydrophilic sequence – IRRPAAKLE – made the protein soluble and allowed crystallization.

• A 13 residue deletion
E V L N E N L L R **F F** V A
in α-casein makes the molecule more soluble.

•***Phenylalanine residues are likely to self-interact and are frequently found at subunit interfaces.***

Amino acid		Transfer free energy kJ/mol	% buried	Chou-Fasman coil index
Phe	F	15.5	48%	0.71
Met	M	14.2	50%	0.58
Ile	I	13	65%	0.66
Leu	L	11.7	41%	0.68
Val	V	10.9	56%	0.62
Cys	C	8.4	47%	1.18
Trp	W	7.9	23%	0.75
Ala	A	6.7	38%	0.7
Thr	T	5	25%	1.07
Gly	G	4.2	37%	1.5
Ser	S	2.5	24%	1.82
Pro	P	-0.8	24%	1.59
Tyr	Y	-2.9	13%	1.06
His	H	-12.5	19%	1.06
Gln	Q	-17.1	6%	0.86
Asn	N	-20.1	10%	1.35
Glu	E	-34.3	20%	1.2
Lys	K	-36.8	4%	0.98
Asp	D	-38.5	15%	1.2
Arg	R	-51.4	0%	1.04

4.3. protein solubility

4.3.4. Protein engineering to increase solubility

4.3.4.3. Primary structure alterations

- A series of point mutations altered the stability and solubility of insulin.

Asn21 is deamidated in an acid solution, resulting in a dimer formation with Gly, Ser, Thr, Asp, His, and Arg.

- Specific sequence changes in proteins from a thermophilic organism show a tendency to replace lysine and glutamic acid with **arginine and aspartic acid** and a preference for the hydrophobic amino acids **Phe, Val and Ile** over Leu, Ala and Met.

- ***Most of these changes occur in α-helical regions and increase the net hydrophobicity of the residue.***

Amino acid		Transfer free energy kJ/mol	% buried	Chou-Fasman coil index
Phe	F	15.5	48%	0.71
Met	M	14.2	50%	0.58
Ile	I	13	65%	0.66
Leu	L	11.7	41%	0.68
Val	V	10.9	56%	0.62
Cys	C	8.4	47%	1.18
Trp	W	7.9	23%	0.75
Ala	A	6.7	38%	0.7
Thr	T	5	25%	1.07
Gly	G	4.2	37%	1.5
Ser	S	2.5	24%	1.82
Pro	P	-0.8	24%	1.59
Tyr	Y	-2.9	13%	1.06
His	H	-12.5	19%	1.06
Gln	Q	-17.1	6%	0.86
Asn	N	-20.1	10%	1.35
Glu	E	-34.3	20%	1.2
Lys	K	-36.8	4%	0.98
Asp	D	-38.5	15%	1.2
Arg	R	-51.4	0%	1.04

4.3. Protein solubility

4.3.4. Protein engineering to increase solubility

4.3.4.4. Post-isolation alterations

• One can alter the solubility of isolated proteins in vitro by coupling to **polyethylene glycol** (Knauf et al., 1988).

4.3.4.5. Designer proteins

A site directed mutagenesis might simply replace a surface **hydrophobic** amino acid with **acidic** residues when aggregation problems arise.

Obviously, the problem of designing soluble proteins is greatly dependent on the ability to predict protein structure.

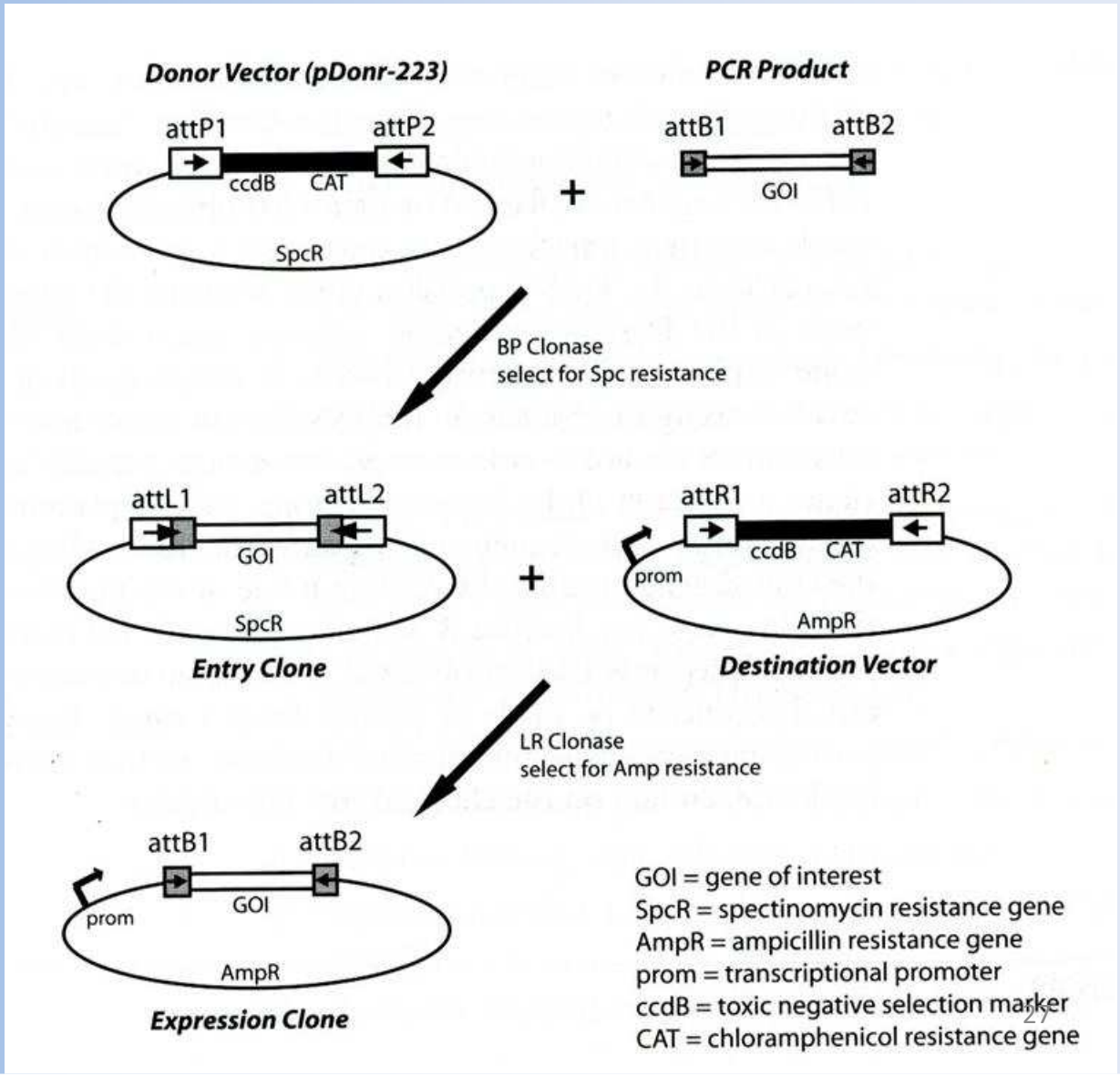
www.expasy.ch

Amino acid		Transfer free energy kJ/mol	% Buried	Chou-Fasman coil index
Phe	F	15,5	48%	0.71
Met	M	14,2	50%	0.58
Ile	I	13	65%	0.66
Leu	L	11,7	41%	0.68
Val	V	10,9	56%	0.62
Cys	C	8,4	47%	1.18
Trp	W	7,9	23%	0.75
Ala	A	6,7	38%	0.7
Thr	T	5	25%	1.07
Gly	G	4,2	37%	1.5
Ser	S	2,5	24%	1.82
Pro	P	-0,8	24%	1.59
Tyr	Y	-2,9	13%	1.06
His	H	-12,5	19%	1.06
Gln	Q	-17,1	6%	0.86
Asn	N	-20,1	10%	1.35
Glu	E	-34,3	20%	1.2
Lys	K	-36,8	4%	0.98
Asp	D	-38,5	15%	1.2
Arg	R	-51,4	0%	1.04

4.4. Gene cloning

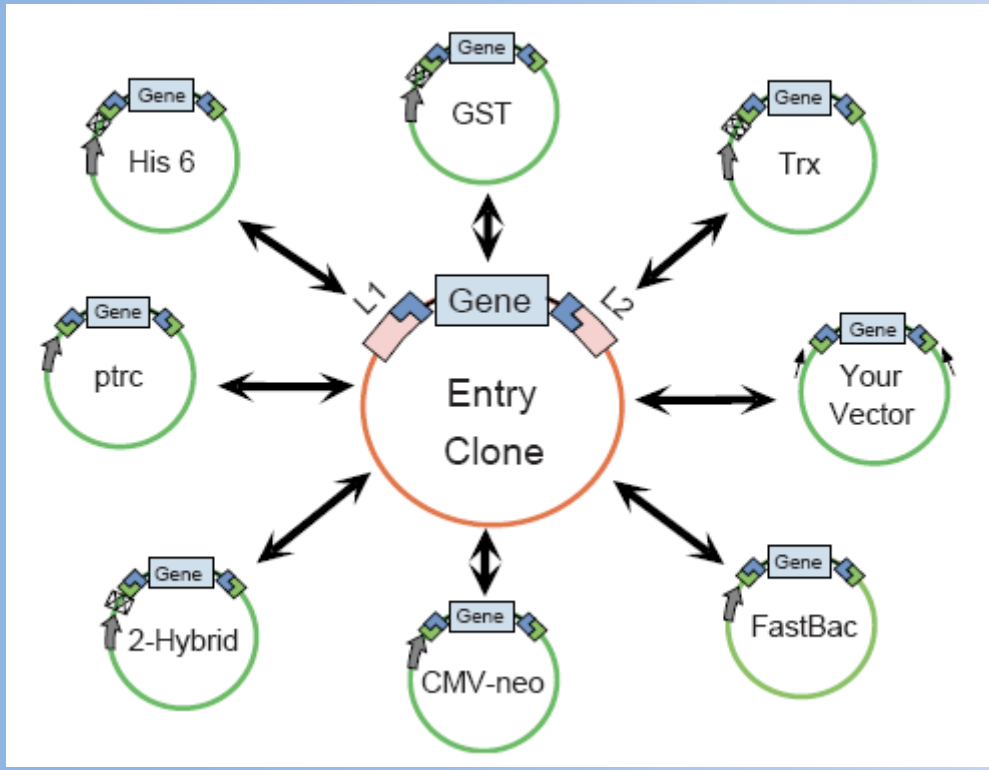
4.4.1. Gateway cloning for protein expression

The protein encoding by ccdB gene interferes with the activity of DNA gyrase and acts to inhibit partitioning of the chromosomal DNA.

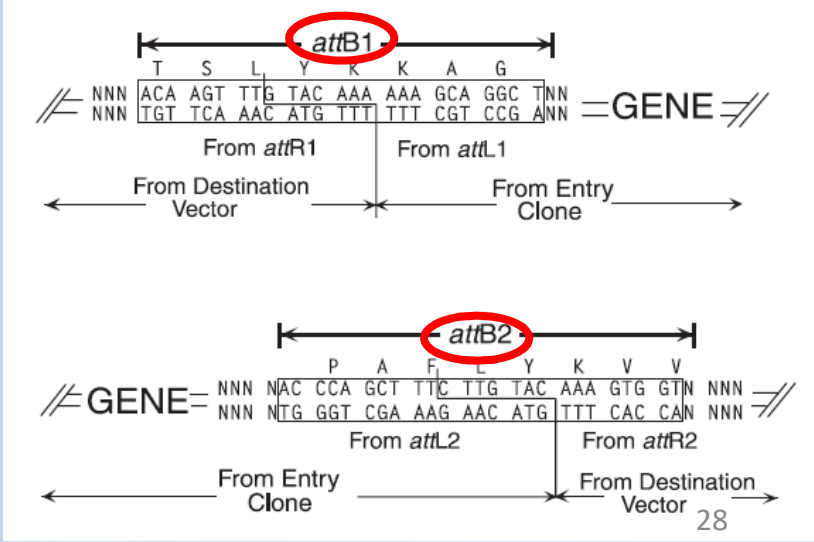


4.4. Gene cloning

4.4.1. Gateway cloning for protein expression



- PCR reaction of the gene containing the terminal *att* sites
- BP reaction of the 1st cloning
- Entry clone – entry vector
- LR reaction of the 2nd cloning
- Destination vector – terminal vector



4.4. Gene cloning

4.4.1. Gateway cloning for protein expression

GOI-stop	Aminoterminal fusions
GOI-nonstop	Aminoterminal and/or carboxyterminal fusions
Kozak-GOI-stop	Aminoterminal fusions or native eukaryotic expression
TEV-GOI-stop	Cleavable aminoterminal fusions
TEV-GOI-Tag	Cleavable aminoterminal fusions with carboxyterminal epitope/purification tag
SD-GOI-stop	Native expression in *E. coli*
Tag-GOI-stop	Aminoterminal tag inside the entry clone

4.4. Gene cloning

4.4.2. Flexi vector cloning

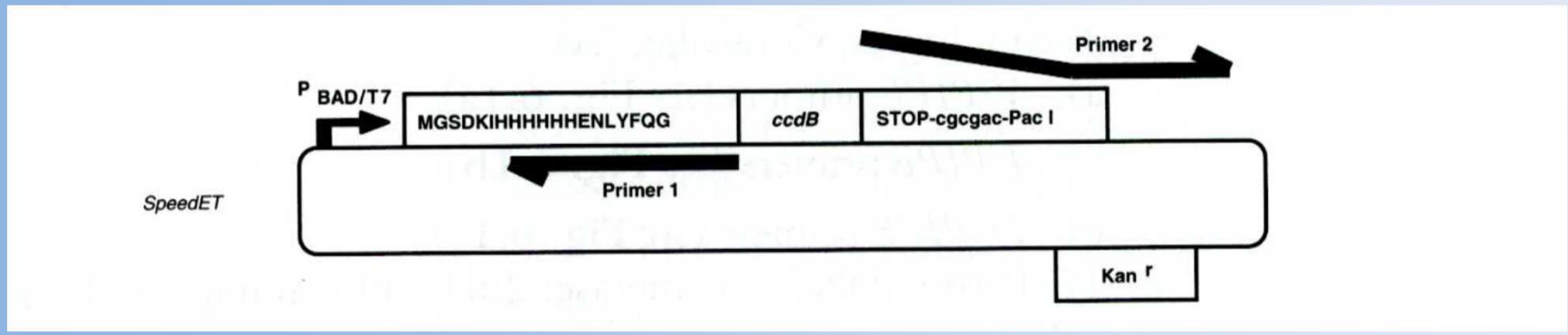
Ligation-dependent cloning method facilitated by selection for the replacement of a toxic gene insert in an acceptor vector.

http://plasmid.hms.harvard.edu

Cloning efficiency:

Human	98.9%
Mouse	98.9%
Rat	98.8%
C. elegans	98.5%
Zebra fish	97.8%
Arabidopsis	97.6%
Yeast	97%

4.4.3. The polymerase primer extension (PIPE)



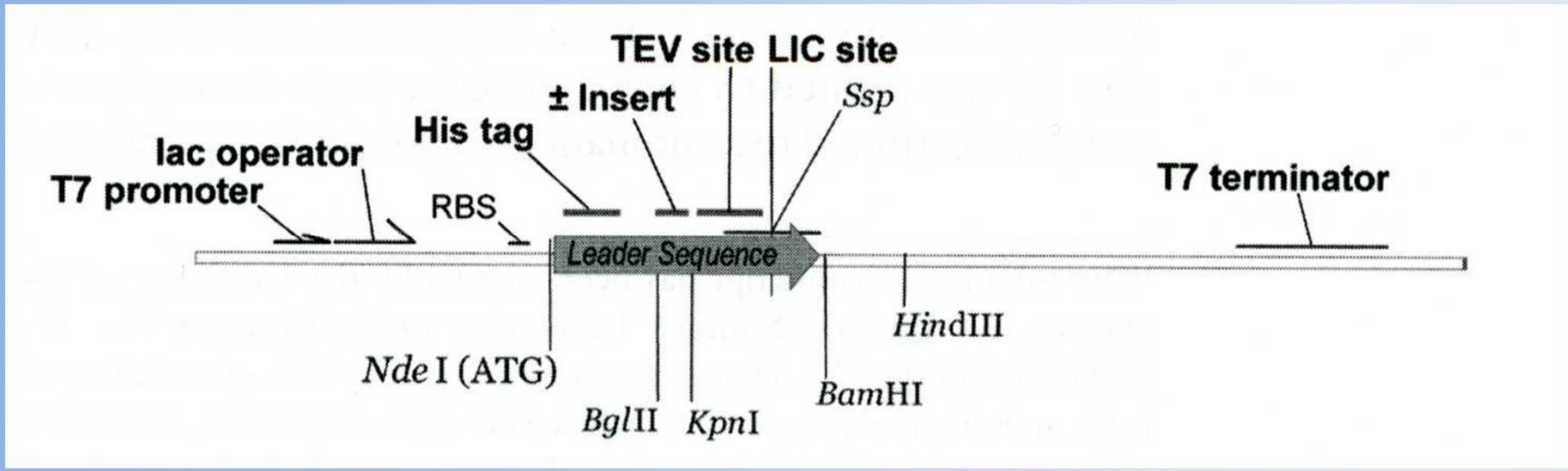
4.4. Gene cloning

4.4.4. In-fusion PCR cloning

http://bioinfo.clontech.com/infusion/

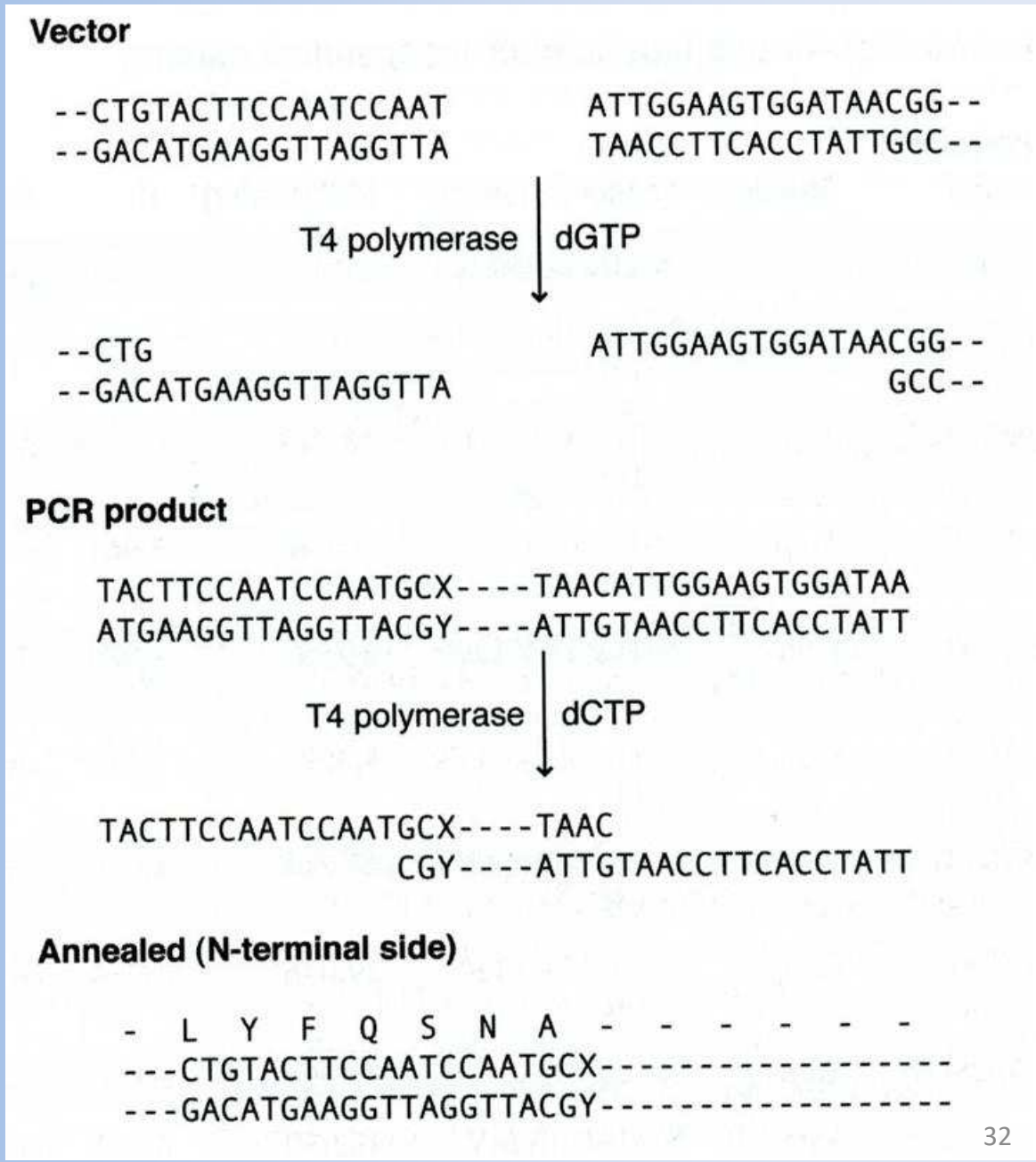
The system is based on an enzyme with proof-reading exonuclease activity that catalyses the joining of DNA duplexes via exposure of complementary single-stranded sequences.

4.4.5. LIC vectors



4.4. Gene cloning

4.4.5. LIC vectors



4.4. Gene cloning

4.4.6. High-throughput cloning and protein expression analysis

Process Workflow

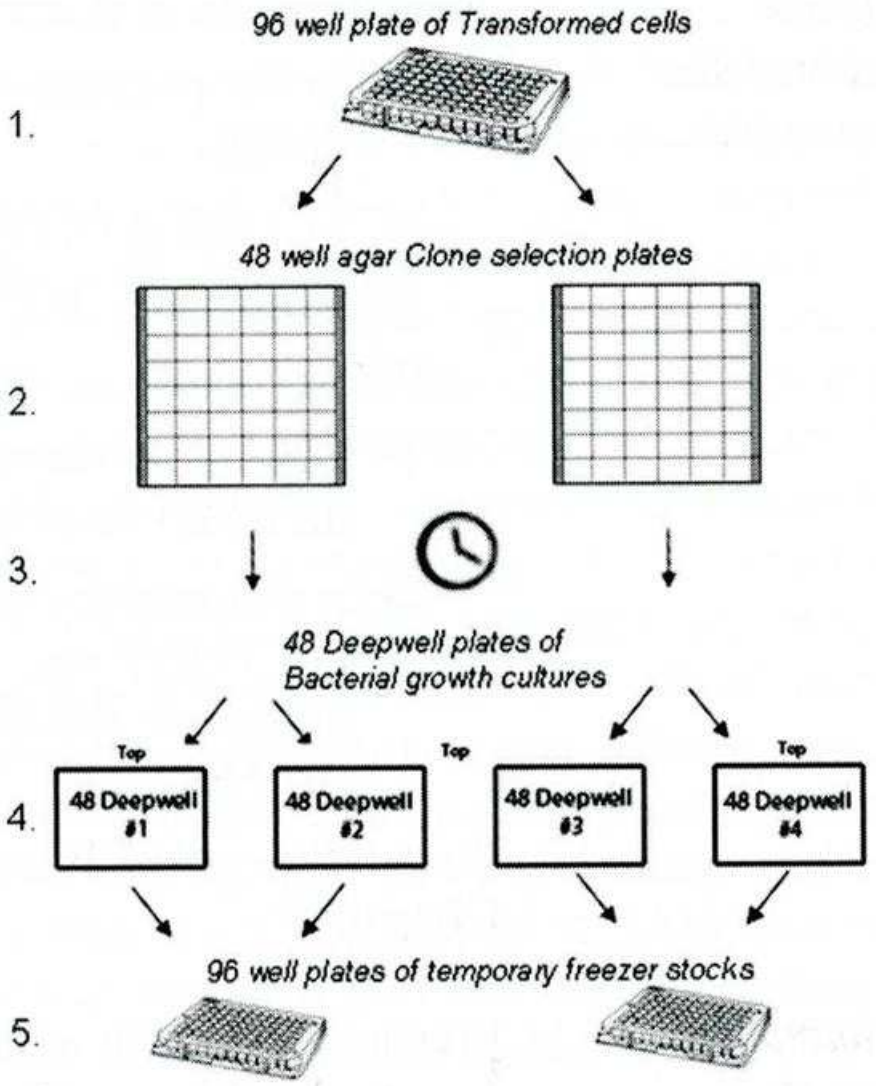
Stage 1:
Vector annealing and cell transformation
(Prepared with Robots)

Stage 2:
Plating for individual clone selection
(Prepared Manually)

Stage 3:
Overnight growth @ 37°C

Stage 4:
Transfer select colonies into Bacterial growth cultures

Stage 5:
Remove aliquot as a temporary freezer stock



4.4. Gene cloning

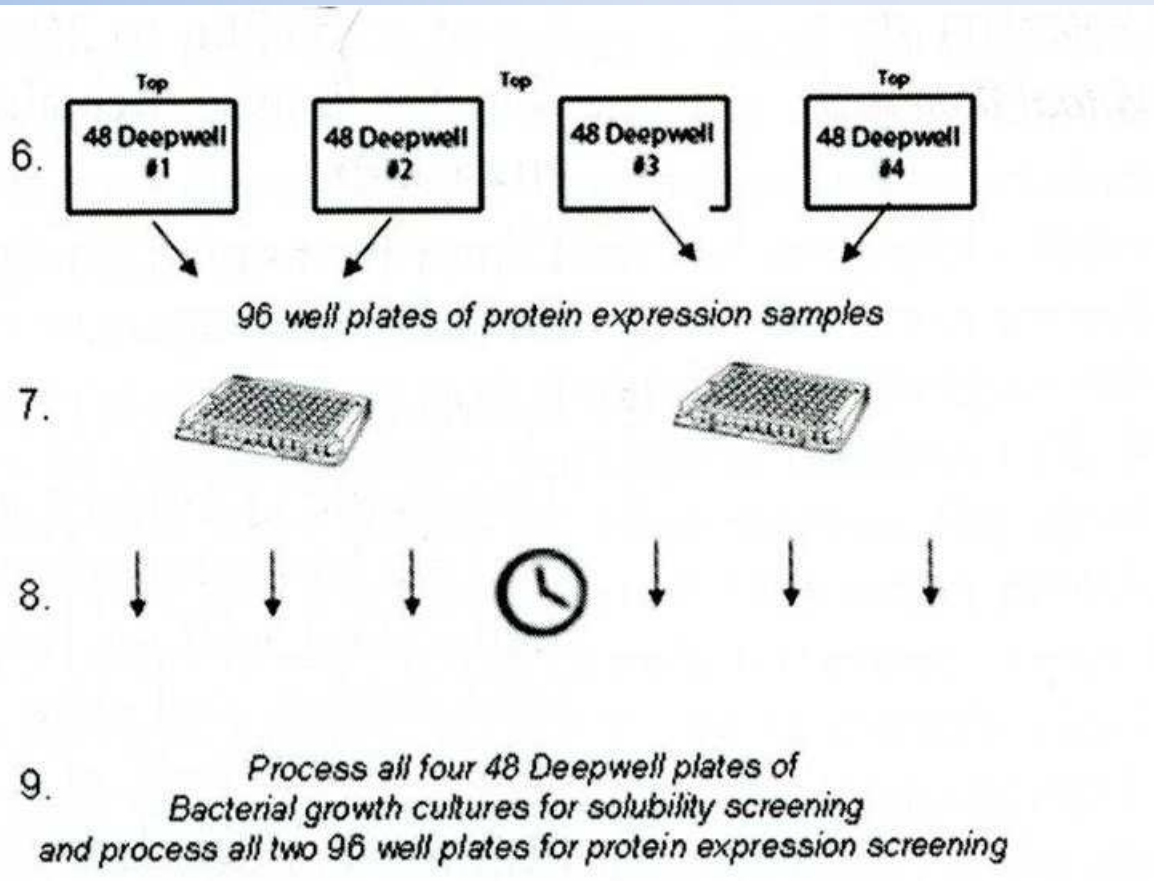
4.4.6. High-throughput cloning and protein expression analysis

Stage 6:
IPTG addition to growth cultures for induction of protein expression

Stage 7:
Aliquot removal for protein expression screening

Stage 8:
Centrifugation of protein expression samples and 48 Deepwell plates of Bacterial growth culture

Stage 9:
Process all plates for expression and solubility screening



4.5. Gene synthesis

Quotation #:

Gene name: CKI1rd

Customer:

Optimized for expression in: *E. coli*
S. cerevisiae(gbpin)

Gene length: 436

Optimization region: 6-422

Analysis conducted by: Jason Zhou, Ph.D

Analysis created: 08/24/2011 01:14:06

QA: James

GenScript
The Biology CRO

Innovation Partner in Drug Discovery!

OptimumGene™ Codon

Op

OptimumGene™ Codon Optimization Analysis

Optimization Parameters

OptimumGene™ algorithm optimizes a variety of parameters that are critical to the efficiency of gene expression, including but not limited to:

- Codon usage bias
- GC content
- CpG dinucleotides content
- mRNA secondary structure
- Cryptic splicing sites
- Premature PolyA sites
- Internal chi sites and ribosomal binding sites
- Negative CpG islands
- RNA instability motif (ARE)
- Repeat sequences (direct repeat, reverse repeat, and Dyad repeat)
- Restriction sites that may interfere with cloning

860 Centennial Ave., P
Tel: 1-732-885-9188 Fax: 1-732-210-0262

4.5. Gene synthesis

Results *E. coli*

1. Codon usage bias adjustment

Frequency of Optimal Codons (FOP)

2. GC Content Adjustment

GC Content Adjustment

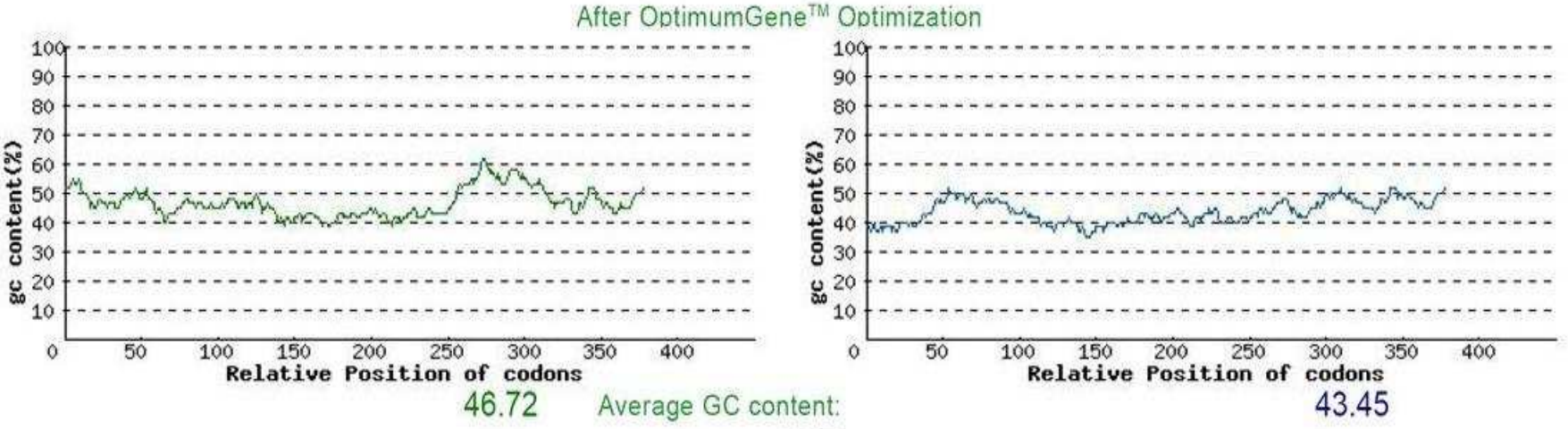


Figure 2. The ideal percentage range of GC content is between 30-70 %. Peaks of %GC content in a 60 bp window have been removed.

4.5. Gene synthesis

3. Restriction Enzymes and CIS-Acting Elements

Restriction Enzymes	Optimized	Original
* Green: filtered sites; Blue: checked sites (not filtered); Red: kept sites.		
NcoI(CCATGG)	1(1)	1(1)
SalI(GTCGAC)	1(431)	1(431)
BamHI(GGATCC)	1(394)	1(394)
XhoI(CTCGAG)	1(363)	1(363)
HindIII(AAGCTT)	1(371)	2(118,371)

CIS-Acting Elements	Optimized	Original
E.coli_RBS(AGGAGG)	0	0
PolyT(TTTTTT)	0	0
PolyA(AAAAAAA)	0	0
Chi_sites(GCTGGTGG)	0	0
T7Cis(ATCTGTT)	0	0

4. Remove Repeat Sequences

After Optimization

Max Direct Repeat: None
Max Inverted Repeat: None
Max Dyad Repeat: None

Before Optimization

Max Direct Repeat: Size:8 Distance:121 Frequency:2
Max Inverted Repeat: None
Max Dyad Repeat: None

4.5. Gene synthesis

5. **Optimized Sequence** (Optimized Sequence Length:436, GC%:46.72)
6. **DNA Alignment (Optimized Region)**

Conclusion

A wide variety of factors regulate and influence gene expression levels, and our OptimumGene™ algorithm takes into consideration as many of them as possible, producing the single gene that can reach the highest possible level of expression.

In this case, the native gene employs tandem rare codons that can reduce the efficiency of translation or even disengage the translational machinery. We changed the codon usage bias in E. coli by upgrading the CAI from 0.61 to 0.74 , and in S. cerevisiae(gbpln) by optimizing the CAI from 0.70 to 0.69 . GC content and unfavorable peaks have been optimized to prolong the half-life of the mRNA. The Stem-Loop structures, which impact ribosomal binding and stability of mRNA, were broken. In addition, our optimization process has screened and successfully modified those negative cis-acting sites as listed in the introduction.

We are honored to deliver the analysis that you requested. We hope that you are pleased with your GenScript OptimumGene™ results.

4.5. Gene synthesis

Comparison of costs

- **1 mutation introduced by QuikChange + recloning**
  - working time: **9 hours** + 9.5 hours
  - total time: **2-3 weeks** + 1 week
  - price: **7,232 CZK** + 2,898 CZK ~ **10,500 CZK**

- **synthesis of 1000 bp gene + recloning**
  - working time: **1 hour** + 9.5 hours
  - total time: **2-5 weeks** + 1 week
  - price: **7,450 CZK** → **16,000 CZK** + 2,898 CZK

    ~ 11,000 - 19,000 CZK

4.5. Gene synthesis

GenScript
The Biology CRO

Your Innovation Partner in Drug Discovery!

Quotation

Radka Dopitova

Masaryk University

Quote Date: 2011-08-24

Valid Through: 2011-11-24

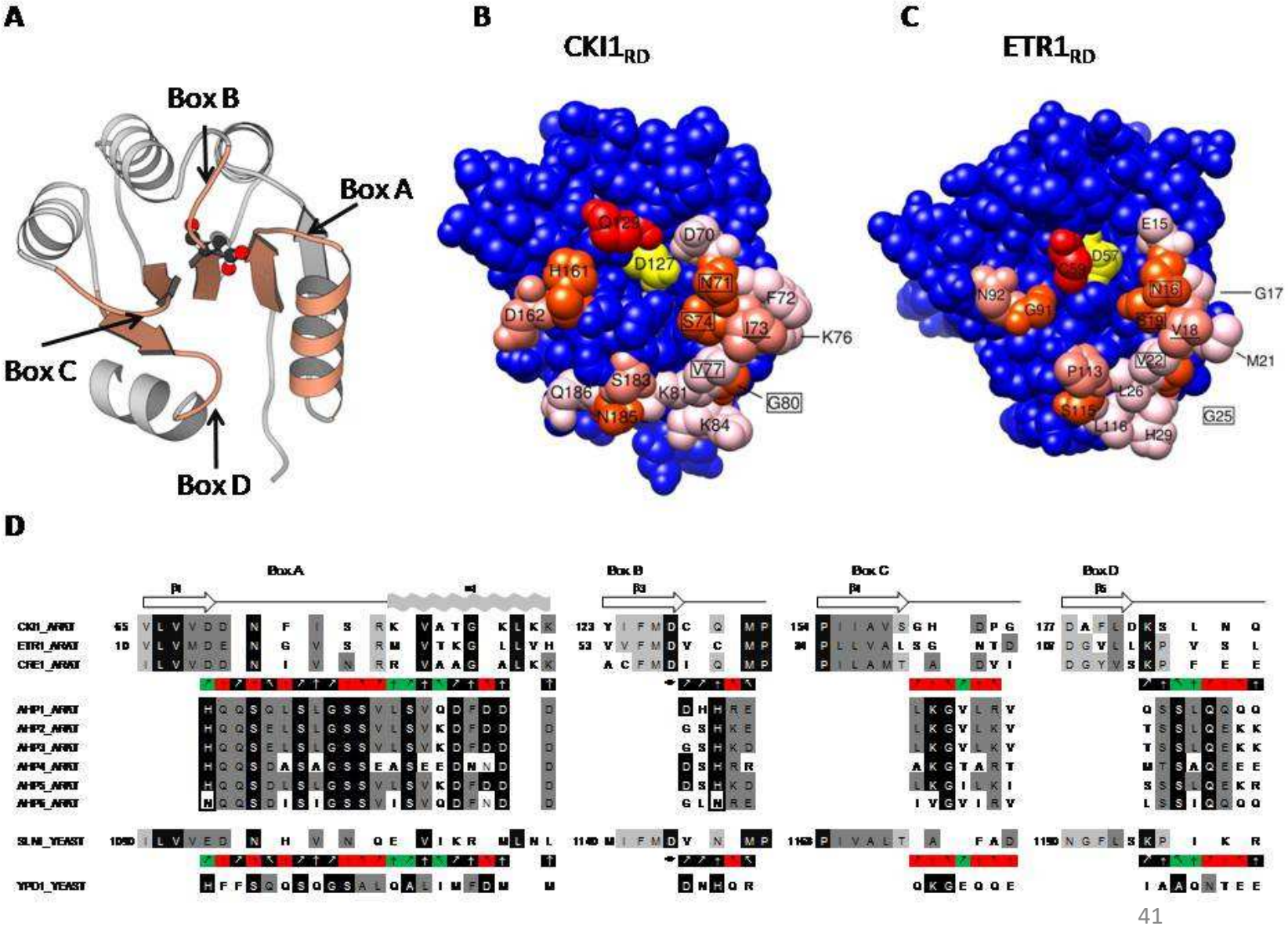
Quote No.	Currency	Estimated Turnaround Time	Terms	Ship Via
985410	United States Dollar ($)		Net 30	FedEx

Quantity	Description	Unit Price	Unit Discount	Extended Price
1	Gene Synthesis: CKI1rd, Len: 436 bp, Vector: pUC57, Cloning site: EcoRV, Quantity: 4 ug	$159.00	$0.00	$159.00
	Subtotal (United States Dollar)	$159.00	$0.00	$159.00
	Estimated Shipping/Handling (United States Dollar)			$90.40
	Total Quote (United States Dollar)			$249.40

Comments:
This is only for reference, not an order confirmation. All payments must be made in United States Dollar. This quotation may not include incidental charges, such as shipping and tax. No sales tax is charged for all international orders. The total charge will be determined at the time the order is placed.

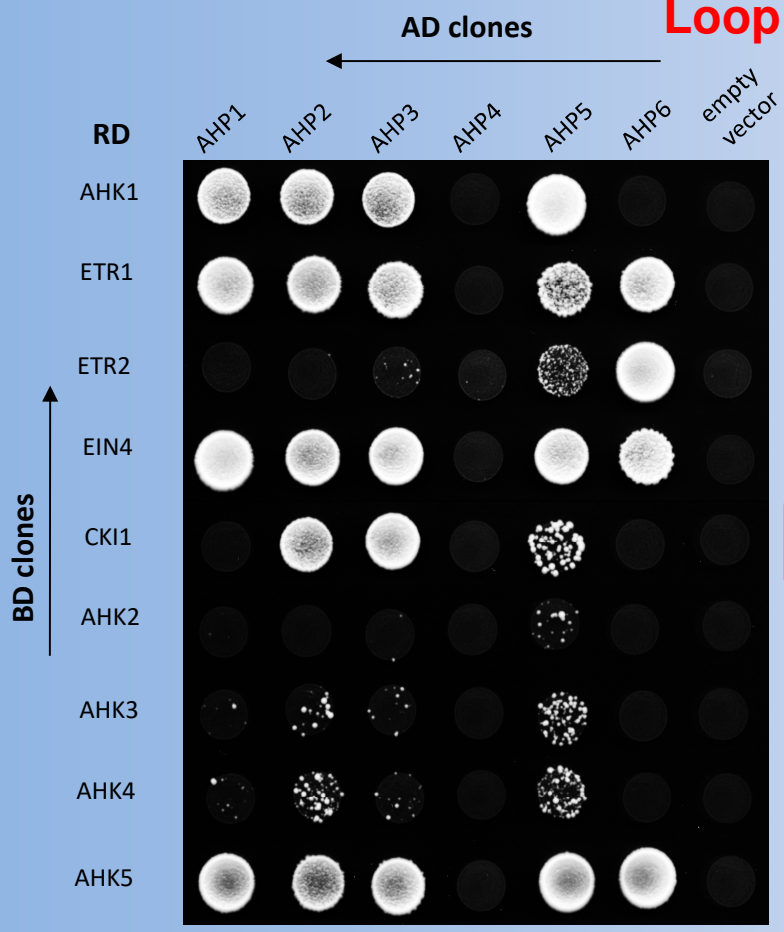
4.5. Gene synthesis

Model of intermolecular contacts.



4.5. Gene synthesis

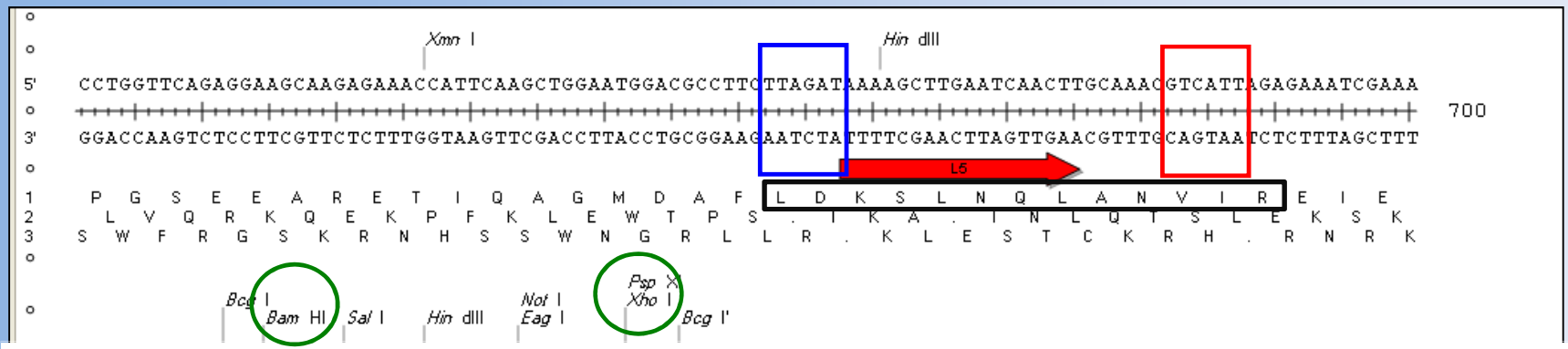
The Y2H results in comparison to aligmnet of RD domains in Loop 5 (Box D)

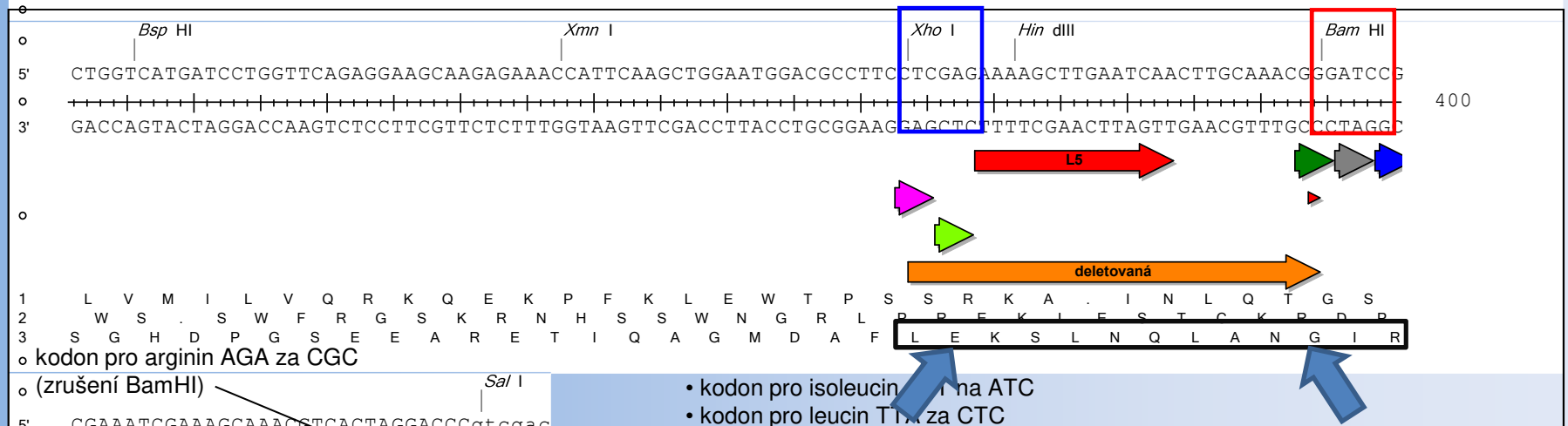


L5 – Box D

P	I	D	R	K	L	M
P	V	S	L	D	N	I
P	V	V	L	R	A	M
P	V	L	L	H	V	M
S	L	N	Q			L
P	F	E	E	E	V	L
P	F	E	A	E	Q	L
P	F	E	E	E	N	L
P	V	T	L	Q	K	L

4.5. Gene synthesis

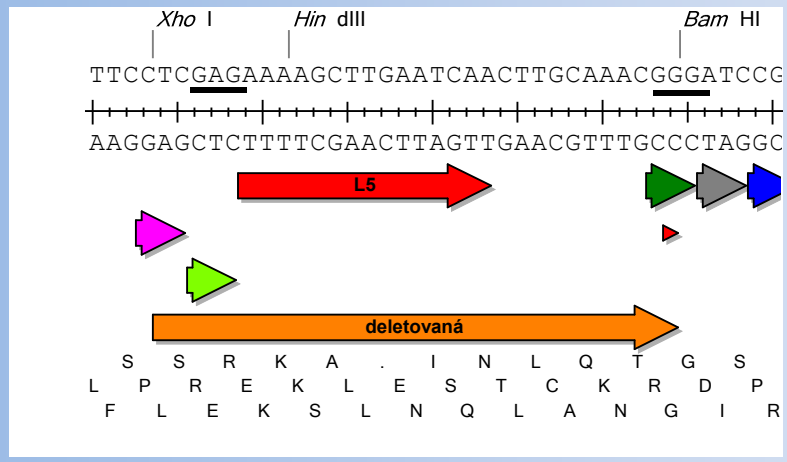




- kodon pro isoleucin [illegible] na ATC
- kodon pro leucin TTA za CTC
- kodon pro valin GTC na GTG
- záměna aspartatu (D) za glutamat (E), kodon GAT na kodon GAG
- záměna valinu za glycin, kodon GTG na kodon GGG
- kodon pro arginin AGA za CGC (zrušení BamHI)

Celkový počet mutovaných nukleotidů: 9
Celkový počet mutovaných kodonů: 6
Celkový počet mutovaných aminokyselin:2

záměna aspartátu (D) za glutamát (E), kodon GAT na kodon GAG
záměna valinu za glycin, kodon GTG na kodon GGG



XhoI

...CTCGAG...
...GAGCTC...

...C...
...GAGCT...

BamHI

...GGATCC...
...CCTAGG...

GATCC...
G...

...C**TCGAT……L5 AHK5…….** GATCC...
...GAGCT**A……L5 AHK5… GT**CTAGG…

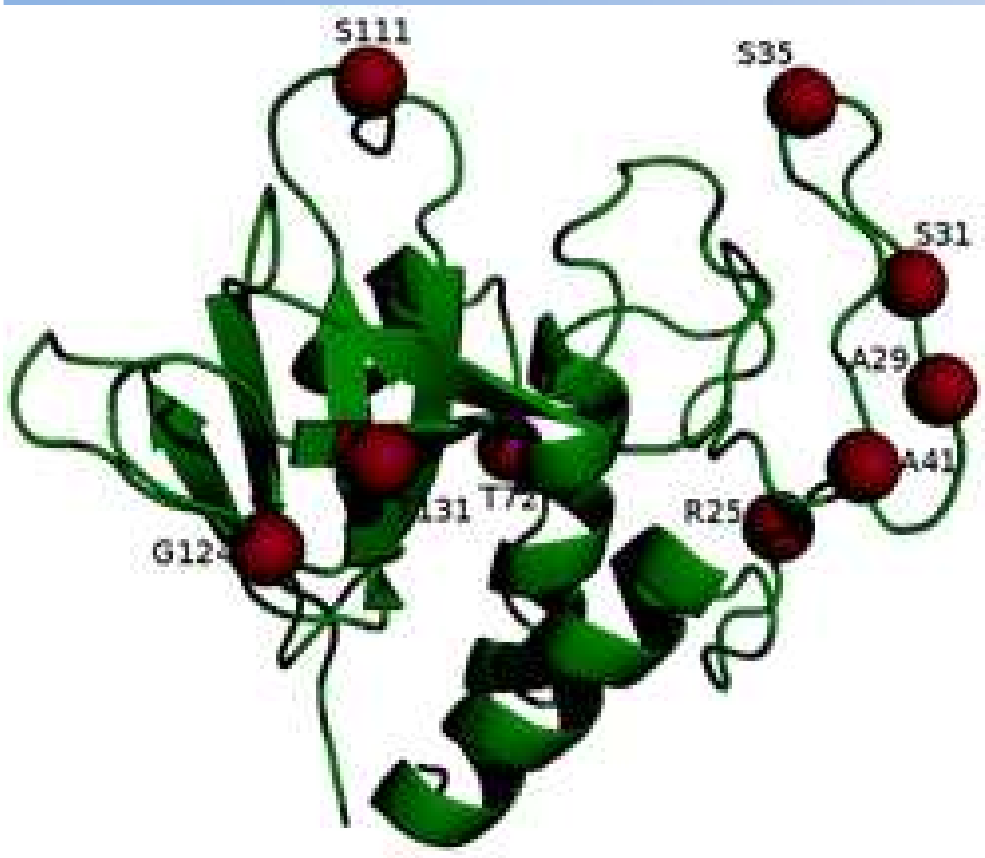
L – D

V – I

Loop exchange

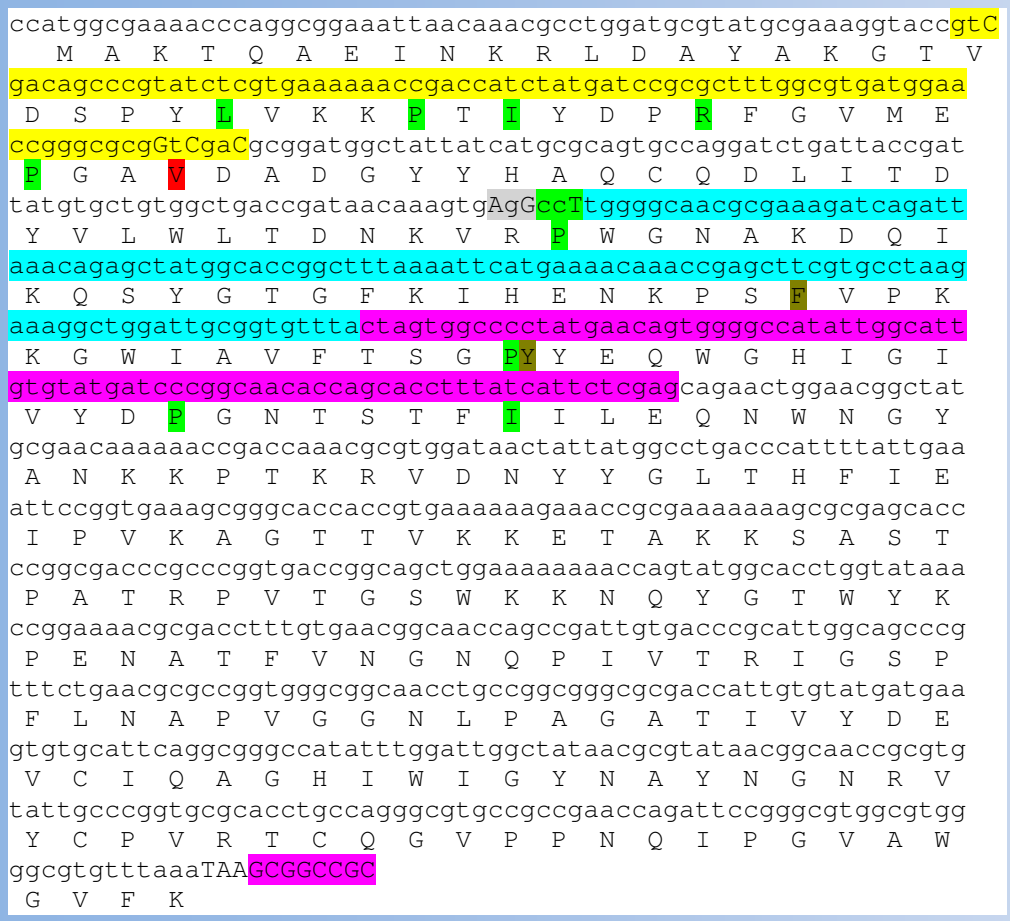
CKI1_ARAT	L	D	K	S	L	N	Q			L	A	N		V		I		R			
Old DNA	TTA	GAT	AAA	AGC	TTG	AAT	CAA			CTT	GCA	AAC	G	T	C	A	T	T	A	G	A
New DNA	CTC	GAG	AAA	AGC	TTG	AAT	CAA			CTT	GCA	AAC	G	G	G	A	T	C	C	G	C
	L	E	K	S	L	N	Q			L	A	N		G		I		R			
AHK2_ARAT			K	P	F	E	E	E	V	L											
AHK3/2_ARAT			K	P	F	E	A	E	Q	L											
CRE1/2_ARAT			K	P	F	E	E	E	N	L											
ETR1_ARAT			K	P	V	S	L	D	N	I											
	TC	GAT	AAA							CTT	GCA	AAC	G	T							
		A	TTT							GAA	CGT	TTG	C	A	C	T	A	G			
	L	D	K							L	A	N		V			I				

CHAP domain – mutagenesis: cloning strategy




```
atggcgaaaacccaggcggaaattaacaaacgcctggatgcgtatgcgaaaggcaccgtg
 M  A  K  T  Q  A  E  I  N  K  R  L  D  A  Y  A  K  G  T  V
gatagcccgtatctcgtgaaaaaaccgaccatctatgatccgcgctttggcgtgatggaa
 D  S  P  Y  L  V  K  K  P  T  I  Y  D  P  R  F  G  V  M  E
ccggggcgcgattgatgcggatggctattatcatgcgcagtgccaggatctgattaccgat
 P  G  A  I  D  A  D  G  Y  Y  H  A  Q  C  Q  D  L  I  T  D
tatgtgctgtggctgaccgataacaaagtgcgccccctggggcaacgcgaaagatcagatt
 Y  V  L  W  L  T  D  N  K  V  R  P  W  G  N  A  K  D  Q  I
aaacagagctatggcaccggctttaaaattcatgaaaacaaaccgagcttcgtgccgaaa
 K  Q  S  Y  G  T  G  F  K  I  H  E  N  K  P  S  F  V  P  K
aaaggctggattgcggtgtttaccagcggccctatgaacagtggggccatattggcatt
 K  G  W  I  A  V  F  T  S  G  P/Y Y  E  Q  W  G  H  I  G  I
gtgtatgatcccggcaacaccagcacctttatcattctggaacagaactggaacggctat
 V  Y  D  P  G  N  T  S  T  F  I  I  L  E  Q  N  W  N  G  Y
gcgaacaaaaaaccgaccaaacgcgtggataactattatggcctgacccatttattgaa
 A  N  K  K  P  T  K  R  V  D  N  Y  Y  G  L  T  H  F  I  E
attccggtgaaagcgggcaccaccgtgaaaaaagaaaccgcgaaaaaaagcgcgagcacc
 I  P  V  K  A  G  T  T  V  K  K  E  T  A  K  K  S  A  S  T
ccggcgacccgcccggtgaccggcagctggaaaaaaaaccagtatggcacctggtataaa
 P  A  T  R  P  V  T  G  S  W  K  K  N  Q  Y  G  T  W  Y  K
ccggaaaacgcgacctttgtgaacggcaaccagccgattgtgacccgcattggcagcccg
 P  E  N  A  T  F  V  N  G  N  Q  P  I  V  T  R  I  G  S  P
tttctgaacgcgccggtgggcggcaacctgccggcgggcgcgaccattgtgtatgatgaa
 F  L  N  A  P  V  G  G  N  L  P  A  G  A  T  I  V  Y  D  E
gtgtgcattcaggcgggccatatttggattggctataacgcgtataacggcaaccgcgtg
 V  C  I  Q  A  G  H  I  W  I  G  Y  N  A  Y  N  G  N  R  V
tattgcccggtgcgcacctgccagggcgtgccgccgaaccagattccgggcgtggcgtgg
 Y  C  P  V  R  T  C  Q  G  V  P  P  N  Q  I  P  G  V  A  W
ggcgtgtttaaa
 G  V  F  K
```

Create cloning sites for the cassettes without changing the restriction site.



SalI-SalI

GTCGAC/GTCGAC

78 nt

StuI-SpeI

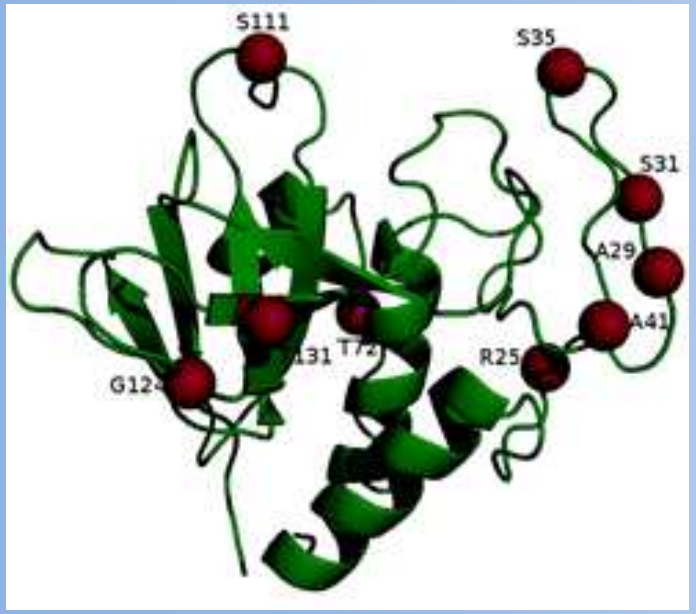
AGGCCT/ACTAGT

117 nt

SpeI-XhoI

ACTAGT/CTCGAG

81 nt



Mutation	Group	Restriction sites
R25L	Red	SalI - SalI
A29P	Red	SalI - SalI
S31K	Red	SalI - SalI
S35R	Red	SalI - SalI
A41P	Red	SalI - SalI
T72P	Blue	StuI - SpeI
T97F	Blue	StuI - SpeI
S111P	Green	SpeI - XhoI
G124P	Green	SpeI - XhoI
T131I	Green	SpeI - XhoI