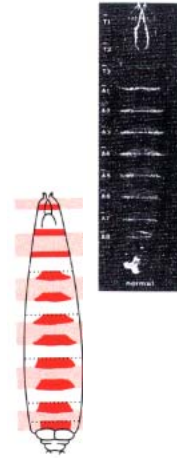


Multi-domain protein example: PAX gene family

- Developmental regulatory genes that encode transcription factors
- Contain a DNA binding domain
- Early expressed during embryogenesis
- Role in morphological boundaries and early regionalization



3

<http://www.gene-regulation.com/info/pax.html>

```

1  mphnsirsgh  gglnglggaf  vngxrplpevv  rqrivdlahq  gvrpcdisrq  lrvshgcvs
61  ilgryyetgs  irpgviggsk  pkvatpkvve  kigdykrqnp  tmfaweirdr  llaegvcdnd
121 tvpsvssinr  iirtkvqqpf  nlpmdscvat  kslspghtli  pssavtppes  pqsdslgsty
181 singllgiaq  pgndnkrkmd  dsdqdsclrs  idsqssssgp  rkhlrtdtfs  qhhlealecp
241 ferqhypeay  aspshtkgeq  glyplpllns  alddgkatlt  ssntplgrnl  sthqtypvva
301 dphspfaikq  etpelsssss  tpsslsssaf  ldlqqvsgg  pagasvppfn  alphaasvyg
361 qftgqallsg  remvgptlpg  ypphiptsgq  gsyassaia  mvagseysgn  ayshtpyssy
421 seawrfpnss  llspyyys  tsrpsappts  atafdhl

```

paired box gene 8 [Mus musculus]

gi|6754990|ref| [[6754990]

[CDART: Conserved Domain Architecture Retrieval Tool](#)

4

	PAM	BLOSUM
Evolutionary model	Explicit evolutionary model	None
Data	Full length MSAs of closely related sequences.	Conserved blocks. i.e., ungapped local MSAs
Bias correction	Trees	Clustering
Multiple substitutions	Markov model: $P^n = (P^1)^n$	Implicitly represented in data (clustering)
Evolutionary distance	Markov model: $P^n = (P^1)^n$	Clustering
Matrices	Transition and log odds scoring matrices	Log odds scoring matrix only.
Parameter n	Distance increases with n	Distance decreases with n
Biophysical properties	Derived indirectly from data	Derived indirectly from data

Comparing PAM and BLOSUM matrices

	PAM	Sequence identity	BLOSUM
	20	83%	
	30		
	60	63%	
	70		
	100	43%	90
	120	38%	80
	160	30%	60
	200	25%	50
	250	20%	45

More divergent ↓

↑ Less divergent