Supplementary Material 2

All double amino acid pair repeats (DAARs), in which the codons have the complementarity to form a hairpin, are surveyed here. The two pairing amino acids are listed in the first and second columns, and results are noted in the other columns.

		(aa1)5(aa2)5: Multiple searches found no repeat, so the search was abandoned.	(aa1aa2)5; For many of these dipeptides, short repeats, such as trimers, are found, but often separated by other amino acids and diverse repeats. Thus, they do not fold into a hairpin.
UUU (Phe)	AAA (Lys)	No (Phe)5(Lys)5 string found	No (FK)5 found
	AAG (Lys)	"	No (FK)5 found
	AGG (Arg)	No (Phe)5(Arg)5 string found	No (FR)5 found
	GGG (Gly)	No (Phe)5(Gly)5 string found	No (FG)5 found
	GAG (Glu)	No (Phe)5(Glu)5 string found	No (FE)5 found
	GAA (Glu)	"	No (FE)5 found
UUC (Phe)	GAA (Glu)	No (Phe)5(Glu)5 string found	No (FE)5 found
	GAG (Glu)	"	No (FE)5 found
	GGG (Gly)	No (Phe)5(Gly)5 string found	No (FG)5 found
	GGA (Gly)	"	No (FG)5 found
UUA (Leu)	UAA (Stop)		
	UAG (Stop)		
	UGA (Stop)		
	UGG (Trp)	No (Leu)5(Trp)5 string found	No (LW)5 found
UUG (Leu)	CAA (Gln)	No (Leu)5(Gln)5 string found	No (LQ)5 found
	CAG (Gln)	"	No (LQ)5 found
	CGG (Arg)	No (Leu)5(Arg)5 string found	No (LR)5 found
	CGA (Arg)	"	No (LR)5 found
UCU (Ser)	AGA (Arg)	No (Ser)5(Arg)5 string found	(SR) repeats use diverse codons, which prevents hairpin
	AGG (Arg)	"	
	GGA (Gly)		(SG) repeats use diverse codons, which prevents hairpin, e.g. X17042.1
	GGG (Gly)		"
UCC (Ser)	GGA (Gly)		"
	GGG (Gly)		"
UCA (Ser)	UGA (Stop)		
	UGG (Trp)		No (SW)5 found
UCG (Ser)	CGA (Arg)		(SR) repeats use diverse codons, which prevents hairpin
	UGA (Stop)		
	CGG (Arg)		(SR) repeats use diverse codons, which prevents hairpin
	UGG (Trp)		No (SW)5 found
UAU (Tyr)	AUA (Ile)		No (YI)5 found; see note for IY later
	AUG (Met)		No (YM)5 found
	GUA (Val)		No (YV)5 found
	GUG (Val)		No (YV)5 found
UAC (Tyr)	GUA (Val)		No (YV)5 found
	GUG (Val)		No (YV)5 found
UGU (Cys)	ACA (Thr)		No (CT)5 found
	GCA (Ala)		No (CA)5 found

	GCG (Ala)	No (CA)5 found
	AUA (Ile)	No (CI)5 found; instead Ile is conservatively replaced with Val, to avoid base pairing: NP_001265373.1, ankyrin repeat LEM domain protein: 632 CVCVCVCVCVCVCLCVCVCV 651; See CV / VC later.
UGC (Cys)	GCA (Ala)	No (CA)5 found
	GCG (Ala)	No (CA)5 found
	GUA (Val)	None.
	GUG (Val)	In the NP_001265373.1 example above, Cys codons are UGU, so see later (VC repeats).
UGG (Trp)	CCA (Pro)	No (WP)5 found. General finding on W repeat: Spaced, short W repeats are found in unusual proteins, e.g., alternate prion, keratin 9, and unnamed proteins.
	UCA (Ser)	No (WS)5 found
	UUA (Leu)	No (WL)5 found
	CUA (Leu)	
	CUG (Leu)	
	UCG (Trp)	No (WW)5 found
CUU (Leu)	AAG (Lys)	No (LK)5 found
	AGG (Arg)	No (LR)5 found
	GGG (Gly)	No (LG)5 found
	GAG (Glu)	No (LE)5 found
CUC (Leu)	GAG (Glu)	No (LE)5 found, as before
	GGG (Gly)	No (LG)5 found, as before
CUA (Leu)	UAG (Stop)	
	UGG (Trp)	No (LW)5 found; same as (WL)5
CUG (Leu)	CAG (Gln)	No (LQ)5 found
	CGG (Arg)	No (LR)5 found
CCU (Pro)	AGG (Arg)	No (PR)5 found; but short PX repeats occur.
	GGG (Gly)	Pro-Gly repeats are found: E3 Ubq ligase HUWE1, all isoforms, X1-13; such as X1 = XP_016884680.1; RNA-binding protein 27, many isoforms, such as X1 = XP_005268523.1. But see Word file describing avoidance of pairable codons.
CCC (Pro)	GGG (Gly)	"
CCA (Pro)	UGG (Trp)	No (PW)5 found; same as (WP)5
CCG (Pro)	CGG (Arg)	No (PR)5 found, as before.
	UGG (Trp)	No (PW)5 found as (WP)5 not found before.
CAU (His)	AUG (Met)	No (HM)5 found.
	GUG (Val)	No (HV)5 found, but there is (HI)5 repeat, e.g. EAW68840.1. Nature chose a conservative replacement, because His codon does NOT pair with Ile codons!!
CAC (His)	GUG (Val)	See above.
CAA (Gln)	UUG (Leu)	No (LQ)5 found, as (QL)5 not found before.
CAG (Gln)	CUG (Leu)	"
	UUG (Leu)	"

CGU (Arg)	ACG (Thr)	No (RT)5 (same as TR) repeat was found, but it retrieved mixture of SRTRHR etc.
	GCG (Ala)	Purest RA repeat is 89-RARARARARATRARRAVQKRA-109 in NP_057691.1 (armadillo repeat X-linked protein), but the RA length is not long enough for hairpin. More interestingly, it used other (synonymous) codons for both R (e.g. AGN) and A.
CGC (Arg)	GCG (Ala)	See above
	GUG (Val)	None.
CGA (Arg)	UCG (Ser)	See SR comment earlier.
	UUG (Leu)	No (LR)5 found earlier.
CGG (Arg)	CCG (Pro)	No (PR)5 found earlier.
AUU (Ile)	AAU (Asn)	No (IN)5 found.
	AGU (Ser)	No (IS)5 found.
	GGU (Gly)	None. Instead GX repeats are found where X is diverse amino acids; example 525-GMGIGVGTGVDAGMGIGVGTG-545 in XP_016886049.1
AUC (Ile)	GAU (Asp)	No ID repeat found. A few have short IX repeat with diverse X.
	GGU (Gly)	No IG repeat found. A few have short IX or GX repeats with diverse X.
AUA (Ile)	UAU (Tyr)	No IY repeat found. A few have short IX repeats with diverse X. Example in XP_016864375.1, putative protein: 109-IYIYIHTYIHICIYIYMYFYIYVY -132
	UGU (Cys)	See previous comment for CI repeat (CV found instead).
AUG (Met)	CAU (His)	HM not found before.
	CGU (Arg)	
ACU (Thr)	AGU (Ser)	No ST or TS repeat found.
	GGU (Gly)	No GT or TG repeat found.
ACC (Thr)	GGU (Gly)	No GT or TG repeat found.
ACA (Thr)	UGU (Cys)	No (CT)5 found before.
ACG (Thr)	CGU (Arg)	No RT /TR found.
	UGU (Cys)	No (CT)5 found before.
AAU (Asn)	AUU (Ile)	None.
	GUU (Val)	NV run not found; instead diverse aliphatic amino acid is found for X in NX runs; example BCL9L protein AAH33257.1 , 62 - NLNMNMNVNMNMNMNLNV- 79. This was found to be a common strategy!
AAC (Asn)	GUU (Val)	See above.
AAA (Lys)	UUU (Phe)	No FK repeat found before.
AAG (Lys)	CUU (Leu)	None.
	UUU (Phe)	No FK repeat found earlier.
AGU (Ser)	ACU (Thr)	None.
	GCU (Ala)	None.

AGC (Ser)	GCU (Ala)	(S)n(A)n with large loop occurs, but they use Ser UCC codon and a mix of GCN for Ala; Example NKX6.1, AH007313.2	None.
	GUU (Val)		None
AGA (Arg)	UCU (Ser)		Previously found: (SR) repeats use diverse codons, which prevents hairpin
	UUU (Phe)		No (FR)5 found previously.
AGG (Arg)	CCU (Pro)		See PR repeat comment earlier.
	UCU (Ser)		See previous comments.
	UUU (Phe)		No (FR)5 found earlier.
	CUU (Leu)		No (LR)5 found earlier.
GUU (Val)	AAC (Asn)		See NV earlier.
	GGC (Gly)		GV repeat, when found, uses GGU codon for Gly and GUG for Val, which do not pair; example: XM_017030560.1, fibroin heavy chain-like, which is full of all sorts of repeat.
	GAC (Asp)		None.
	AGC (Ser)		None, as before.
	GGU (Gly)		See GV repeat above.
	GAU (Asp)		None.
	AGU (Ser)		None, as before.
GUC (Val)	GAC (Asp)		None, as before.
	GAU (Asp)		None, as before.
	GGC (Gly)		See above.
	GGU (Gly)		See above.
GUA (Val)	UAC (Tyr)		None, as before.
	UGC (Cys)		None, as before.
	UAU (Tyr)		None, as before.
	UGU (Cys)		None, as before.
GUG (Val)	CAC (His)		See earlier.
	UAC (Tyr)		None, as before.
	CGC (Arg)		None, as before.
	UGC (Cys)		None, as before.
	CAU (His)		See HV earlier.
	UAU (Tyr)		No YV repeat found earlier.
	CGU (Arg)		No RV found earlier.
	UGU (Cys)		Rare. Example NP_001265373.1 mentioned before. Has the right codon pair for hairpin. However, this hairpin is just before the translation stop codon. So, translation will stop anyway, and ribosome slowing down may actually help. Moreover, in a long GU / UG run, most G:U bonds do not actually form; so this hairpin is also unstable (only - 6 kcal by Mfold), and may not form.
GCU (Ala)	AGC (Ser)	No A5X5Y5 stretch contains hairpinable GCU and AGC/AGU	None; as with SA before.
	AGU (Ser)		As above.
	GGC (Gly)		A few AG runs are found, some use GCU for Ala, but almost exclusively GGG for Gly, so hairpin cannot form.
	GGU (Gly)		See above.
GCC (Ala)	GGC (Gly)		See above, Very few Ala codons in these repeats are GCC, they are mostly GCU
	GGU (Gly)		See above.
GCA (Ala)	UGC (Cys)		No repeat found before.

	UGU (Cys)	No repeat found before.
GCG (Ala)	CGC (Arg)	Not found. See comment before.
	CGU (Arg)	Not found. See comment before.
	UGC (Cys)	Not found, as before.
	UGU (Cys)	Not found, as before.
GAU (Asp)	AUC (Ile)	ID repeat not found before.
	AUU (Ile)	ID repeat not found before.
	GUC (Val)	None, as before.
	GUU (Val)	None, as before.
GAC (Asp)	GUC (Val)	None.
	GUU (Val)	None.
GAA (Glu)	UUC (Phe)	No FE repeat found before.
	UUU (Phe)	No FE repeat found before.
GAG (Glu)	CUC (Leu)	No LE repeat found before.
	CUU (Leu)	None.
	UUC (Phe)	No FE repeat found before.
	UUU (Phe)	No FE repeat found before.
GGU (Gly)	ACC (Thr)	No TG repeat found before.
	ACU (Thr)	None.
	GCC (Ala)	See AG run before (synonymous codon used that ruled out base pairing).
	GCU (Ala)	See AG run before.
	AUC (Ile)	None; see previous comment on IG repeat.
	AUU (Ile)	None; see previous comment on IG repeat.
	GUC (Val)	None; see previous comment on VG repeat.
	GUU (Val)	None; see previous comment on VG repeat.
GGC (Gly)	GCC (Ala)	None; see AG repeat before.
	GUC (Val)	None; see before.
	GUU (Val)	None; see before.
GGA (Gly)	UCC (Ser)	None; see previous comments.
	UCU (Ser)	None; see previous comments.
	UUC (Phe)	None; see previous comments.
	UUU (Phe)	None; see previous comments.
GGG (Gly)	CCC (Pro)	None; see previous comments.
	CCU (Pro)	None; see previous comments.
	CUC (Leu)	None; see previous comments.
	CUU (Leu)	None; see previous comments.
	UCC (Ser)	None; see previous comments.
	UCU (Ser)	None; see previous comments.
	UUC (Phe)	None; see previous comments.
	UUU (Phe)	None; see previous comments.