

# ANALYSIS OF COLLAGEN TYPE I CHAINS

SORANA-DANIELA BOLBOACĂ AND LORENTZ JÄNTSCHI

Collagen type I, consists of a heterotrimer of two  $\alpha 1(I)$  and one  $\alpha 2(I)$  chains, is the most common form of fibrillar collagen, being a major constituent of bone and skin. The research presents a correlation analysis of amino acids within and between collagen types I chains in the same specie and in different species. The collagen type I chains from the following species were included into analysis: *Rattus norvegicus* (Orjel et al., 2006), *Bos taurus* (Shirai et al., 1998), *Danio rerio* (Howden, 2007), *Canis lupus* (Lowe et al., 2003), and *Homo sapiens* (Strausberg et al., 2002). The correlation obtained were analyzed in accordance with the distribution of amino acids in the collagen type I chains [1],[2]. A perfect correlation ( $r = 1$ ) was obtained between species on the same  $\alpha$  chain for cysteine, tryptophan, tyrosine, and lysine. The highest correlation on different collagen type I chains ( $r = 0.763$ ) of the same species was obtained for *Canis lupus*.

Acknowledgments: CNCSIS Romania supports the research through project AT93/2007.

## REFERENCES

- [1] Bolboacă S, Jäntschi L. Amino Acids Sequences Analysis on Collagen. *BUSAMV - Animal Science and Biotechnologies* 2007;63-64:311-316.
- [2] Bolboacă S, Jäntschi L. A Structural Informatics Study on Collagen. *Chem Biol Drug Des* 2008;71(2):173-179.

"IULIU HATIEGANU" UNIVERSITY OF MEDICINE AND PHARMACY CLUJ-NAPOCA,  
13 EMIL ISAC STREET, 400023 CLUJ-NAPOCA, ROMANIA.

*E-mail address:* [sbolboaca@umfcluj.ro](mailto:sbolboaca@umfcluj.ro)

*URL:* <http://sorana.academicdirect.ro>

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA, ROMANIA, 15 CONSTANTIN DAICOVICIU STREET, 400020 CLUJ-NAPOCA, ROMANIA.

*E-mail address:* [lori@j.academicdirect.org](mailto:lori@j.academicdirect.org)

*URL:* <http://lori.academicdirect.org>

---

1991 *Mathematics Subject Classification*. 62H20 (Measures of association); 92C15 (Developmental biology, pattern formation); 62P10 (Applications to biology and medical sciences).

*Key words and phrases.* Correlation; Amino Acids Sequence; Chains of Type I Collagen.

# CORRELATION ANALYSIS ON COLLAGEN TYPE I CHAINS



# OUTLINE

---

- **BACKGROUND**
- **AIM**
- **COLLAGEN TYPE I - SPECIES**
- **METHOD OF INVESTIGATION**
- **RESULTS**
- **CONCLUSIONS**
- **AKNOWLEDGEMENTS**

# BACKGROUND: COLLAGEN

---

- The main protein of connective tissues (fibrous protein, inextensible, which can be found at the level of connective tissues)
- Twenty-eight types of collagens known (Veit et al., 2006)
- Use: gelatin (food, pharmaceutical, cosmetic, and photography industries)

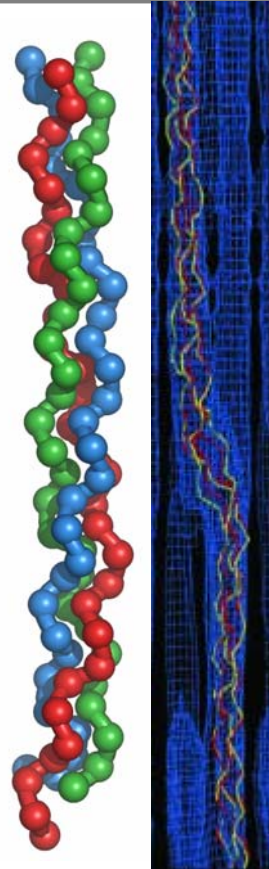
## BACKGROUND: TYPE I COLLAGEN

- Most abundant collagen of the human body
  - Main component of bone (organic part of bones and teeth)
  - Scar tissue & Skin
  - Tendons
  - Artery walls
  - Endomysium of myofibrils
  - Fibrocartilage
- 2  $\alpha 1$  and 1  $\alpha 2$  chains
- COL1A1 (17q21.33) & COL1A2 (7q21.3)
  - osteogenesis imperfecta (Lee et al., 2006)
  - osteoporosis (Ralston et al., 2006)
  - Ehlers-Danlos syndrome (Pollitt et al., 2006)
  - Caffey disease (Gensure et al., 2005),
  - intracranial aneurysms (Yoneyama et al., 2004)
  - bone metastasis (Fukumitsu et al., 2003)

# BACKGROUND: TYPE I COLLAGEN STRUCTURE

(Orgel et al., 2006):

- Amino acids sequence in  $\alpha$  chain (0.84 nm)
- Triple helix ( $\leftrightarrow$  10 nm;  $/ \sim$  1.1 nm)
- Collagen molecule ( $\sim$  298.8 nm)
- Collagen sub-fibril ( $\leftrightarrow$   $\sim$ 67.8 nm &  $\updownarrow \sim$  2.7 nm &  $\leftrightarrow$   $\sim$ 3.9 nm)



## AIM

---

- Amino acids distribution in type I collagen chains ( $\alpha_1$  and  $\alpha_2$ ) on species
- Identify regularities within and between chains and/or species
  - Similarity analysis
  - Rank correlation analysis
  - Autocorrelation analysis

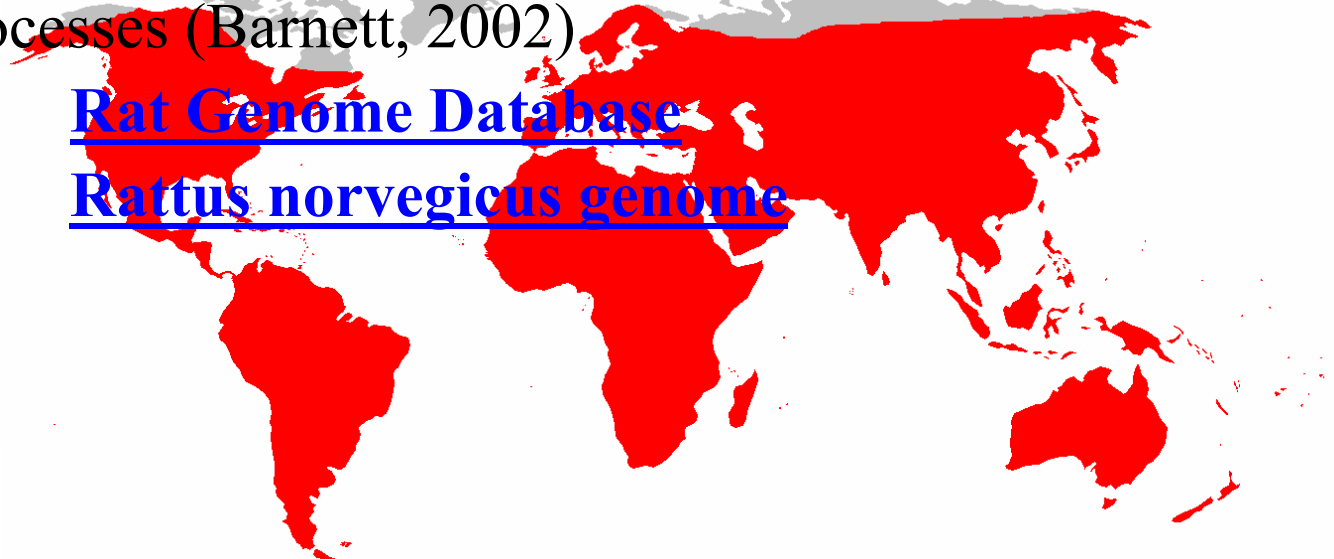
## COLLAGEN TYPE I – INVESTIGATED SPECIES

*Rattus norvegicus* (Orjel at al., 2006):

- Uses in science: since 1895 (Clark University, Massachusetts (United States) studied the effects of diet).
- Experimental studies: understanding genetic - diseases – effects of drugs - psychological studies of learning and other mental processes (Barnett, 2002)



[Rat Genome Database](#)  
[Rattus norvegicus genome](#)





## COLLAGEN TYPE I – INVESTIGATED SPECIES

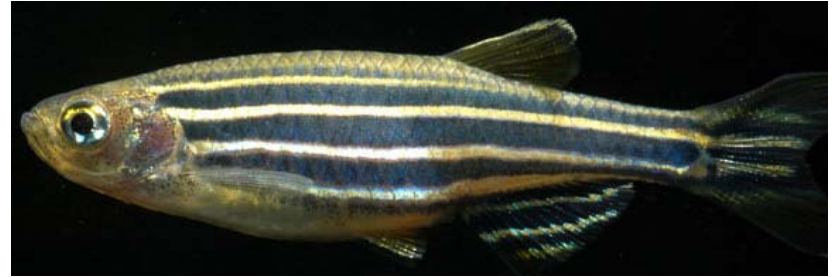
### *Bos taurus* (Shirai et al., 1998)



- o Wild ancestors: northern Africa, Europe, and southern Asia (Nowak, 1997).
- o Economic Importance for Humans:
  - o Positive: dairy products, medicine, glue, soap, leather
  - o Negative: Bovine Spongiform Encephalopathy (Mad Cow Disease)
- o [Bos taurus genome](#)

## COLLAGEN TYPE I – INVESTIGATED SPECIES

*Danio rerio* (Howden, 2007)



- o It is native to the streams of the southeastern Himalayan region (Mayden et al., 2007)
- o Useful model organism for studies of vertebrate development and gene function (George Streisinger - University of Oregon): genetic screens
  - o [The Zebrafish Model Organism Database](#)
  - o [Zebrafish International Resource Center](#)
  - o [FishMap: Zebrafish Genomics Knowledge database](#)

## COLLAGEN TYPE I – INVESTIGATED SPECIES

*Canis lupus* (Lowe et al., 2003)

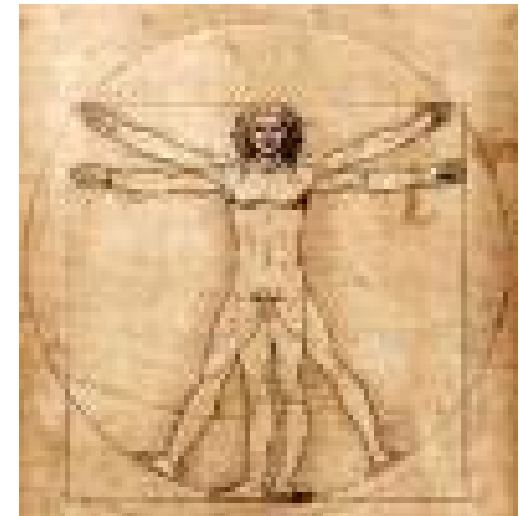


- o Originating during the Late Pleistocene around 300,000 years ago (Nowak, 1992).
- o DNA sequencing and genetic drift studies indicate that the gray wolf shares a common ancestry with the domestic dog, (*Canis lupus familiaris*) and might be its ancestor (Lindblad-Toh et al. 2005).

# COLLAGEN TYPE I – INVESTIGATED SPECIES

*Homo sapiens* (Strausberg et al., 2002)

- o "sapiens" = wise or intelligent
- o has lived from about 250,000
- o publication of name: 1758 (H. sapiens sapiens)
  
- o [Homo sapiens Genome](#)
- o [Ensembl Human Genome](#)
- o [Homo sapiens genome view](#)
- o [KEGG GENOME](#)



## COLLAGEN TYPE I – CHAINS

---

---

- o National Center for Biotechnology Information  
<http://www.ncbi.nlm.nih.gov/>

# COLLAGEN TYPE I – aa DISTRIBUTION

<i>Amino acid (aa)</i>		<i>Homo sapiens</i>		<i>Bos taurus</i>		<i>Canis lupus</i>		<i>Rattus norvegicus</i>		<i>Danio rerio</i>	
Name	Abb	HS_α1	HS_α2	BT_α1	BT_α2	CL_α1	CL_α2	TN_α1	RN_α2	DR_α1	DR_α2
Alanine	A	115	130	143	126	138	123	125	108	162	137
Cysteine	C	10	9	18	9	18	9	0	0	18	8
Aspartate	D	41	43	64	43	64	43	34	23	62	46
Glutamate	E	57	66	76	64	74	65	54	48	82	56
Phenylalanine	F	14	22	24	23	25	21	13	14	28	21
Glycine	G	<b>329</b>	<b>381</b>	<b>389</b>	<b>380</b>	<b>390</b>	<b>381</b>	<b>344</b>	<b>345</b>	<b>382</b>	<b>382</b>
Histidine	H	3	15	9	12	8	12	3	7	10	11
Isoleucine	I	10	32	25	35	26	34	8	19	36	30
Lysine	K	38	50	57	50	56	50	35	21	58	50
Leucine	L	29	61	50	60	47	59	21	34	37	57
Methionine	M	8	10	13	9	15	11	8	5	25	17
Asparagine	N	14	41	29	43	31	42	12	21	35	44
Proline	P	230	231	279	236	278	235	126	113	235	223
Glutamine	Q	30	33	51	36	50	35	25	23	40	37
Arginine	R	51	72	70	73	71	73	52	55	68	72
Serine	S	35	52	58	54	58	52	39	27	62	67
Threonine	T	23	42	44	43	47	51	16	21	53	40
Valine	V	28	55	42	50	42	50	18	39	39	35
Tryptophan	W	1	5	6	5	6	5	0	0	6	5
Tyrosine	Y	3	16	16	13	16	15	5	1	9	14
Unspecified or unknown	X	0	0	0	0	0	0	<b>116</b>	<b>102</b>	0	0
<i>Total</i>		<i>1069</i>	<i>1366</i>	<i>1463</i>	<i>1364</i>	<i>1460</i>	<i>1366</i>	<i>1054</i>	<i>1026</i>	<i>1447</i>	<i>1352</i>

# METHOD: RANK CORRELATION

- o Step 1: matrix representation
  - o Columns: the amino acid of interest (SP\_α1/2\_Z)
  - o Rows: the number of amino acid in the chains (from 0 - when the amino acid was not present on the investigated specie to 390 – glycine on *Canis lupus* - α1 chain).
- o Step 2: calculation of Spearman rank correlation (the rank of each amino acid was correlated with the rank of all other amino acids on the chain for the same specie)

# METHOD: AUTOCORRELATION

---

- o between adjacent entries (an autocorrelation by order  $k = 1$ ).
- o The autocorrelation with an offset of 1 correlate the data set  $\{aa_2, aa_3, aa_4, aa_5, \dots, aa_n\}$  with the data set  $\{aa_1, aa_2, aa_3, aa_4, \dots, aa_{n-1}\}$ .



# METHOD: IMPLEMENTATION

---

- Distribution of aa on type I collagen chains on species: <http://l.academicdirect.org/Agriculture/Colagen/Chains/>
- aa matrix on specie
- Rank correlation matrix
- Autocorrelation matrix

## RESULTS: RANK CORRELATION WITHIN

- Min:
  - 0.2789: DR $\alpha$ 1 L (37 leucine on the chain) - Y (9 tyrosine on the chain)
  - 0.4905: BT $\alpha$ 1 V (42 valine on the chain) - Y (16 tyrosine on the chain) –when all amino acids with appearance less than 10 are excluded
- Max (1.0000):
  - RN $\alpha$ 1 V (18 valine on the chain) - H (3 histidine on the chain)
  - HS $\alpha$ 1 S (35 serine on the chain) - Y (3 tyrosine on the chain)

## RESULTS: RANK CORRELATION WITHIN

n	BT $\alpha$ 1	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W
143	A	1																		
18	C	0.7845	1																	
64	D	0.9575	0.8506	1																
76	E	0.9753	0.8721	0.9876	1															
24	F	0.9431	0.8855	0.8975	0.9582	1														
389	G	0.9896	0.7207	0.9631	0.9750	0.9190	1													
9	H	0.9247	0.8586	0.8998	0.6664	0.8966	0.8874	1												
25	I	0.9779	0.8616	0.7766	0.8682	0.9571	0.9601	0.7540	1											
57	K	0.9737	0.8988	0.9900	0.9910	0.9891	0.9764	0.8965	0.9557	1										
50	L	0.9090	0.9108	0.9861	0.9740	0.8792	0.9296	0.5726	0.7663	0.9737	1									
13	M	0.9619	0.6615	0.9050	0.9226	0.9668	0.9395	0.7646	0.9893	0.9483	0.7925	1								
29	N	0.9247	0.8922	0.8851	0.9492	0.9892	0.9042	0.8184	0.9260	0.9645	0.8667	0.9177	1							
279	P	0.9897	0.8193	0.9651	0.9890	0.9607	0.9971	0.9257	0.9773	0.9837	0.9488	0.9477	0.9355	1						
51	Q	0.9451	0.9036	0.9881	0.9837	0.9727	0.9574	0.8087	0.8996	0.9904	0.9913	0.9475	0.9527	0.9741	1					
70	R	0.9828	0.8271	0.9927	0.9929	0.9750	0.9843	0.9104	0.9384	0.9978	0.9779	0.9685	0.9509	0.9865	0.9917	1				
58	S	0.9638	0.8935	0.9792	0.9651	0.9614	0.9726	0.8393	0.9026	0.9778	0.9790	0.9482	0.9379	0.9634	0.9848	0.9776	1			
44	T	0.9286	0.9003	0.9545	0.9669	0.9614	0.9361	0.9379	0.8675	0.9509	0.9526	0.8762	0.9604	0.9486	0.9628	0.9558	0.9878	1		
42	V	0.9096	0.8569	0.9825	0.9918	0.9459	0.9256	0.8818	0.8347	0.9755	0.9808	0.6875	0.9487	0.9492	0.9877	0.9777	0.9871	0.9868	1	
6	W	0.5094	0.9037	0.8812	0.6770	0.6387	0.5186	0.5822	0.4973	0.5921	0.8151	0.7369	0.5311	0.5932	0.5009	0.7379	0.9302	0.6073	0.9339	1
16	Y	0.8815	0.4957	0.6340	0.6148	0.7488	0.8808	0.9921	0.7474	0.7852	0.5243	0.7537	0.6409	0.8575	0.6682	0.8568	0.7550	0.6500	0.4905	0.5747

## RESULTS: RANK CORRELATION WITHIN

n	RNα1	A	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	Y
125	A	1																	
34	D	0.9807	1																
54	E	0.9956	0.9934	1															
13	F	0.9643	0.9890	0.9775	1														
344	G	0.9972	0.9905	0.9956	0.9852	1													
3	H	0.8849	0.9966	0.6953	0.9000	0.9457	1												
8	I	0.9892	0.8710	0.9214	0.9503	0.9801	0.8491	1											
35	K	0.9751	0.9826	0.9829	0.9713	0.9918	0.9996	0.9041	1										
21	L	0.9575	0.9803	0.9911	0.9721	0.9880	0.9998	0.8386	0.9802	1									
8	M	0.8214	0.9630	0.9215	0.9079	0.8629	0.8074	0.7669	0.9527	0.9528	1								
12	N	0.8953	0.9668	0.9366	0.9634	0.9297	0.9789	0.8034	0.9156	0.9179	0.9189	1							
126	P	0.9924	0.9733	0.9938	0.9897	0.9971	0.8567	0.8435	0.9857	0.9663	0.9830	0.9715	1						
25	Q	0.9734	0.9754	0.9900	0.9679	0.9925	0.6231	0.8689	0.9908	0.9845	0.9601	0.9264	0.9779	1					
52	R	0.9862	0.9826	0.9922	0.9844	0.9978	0.9593	0.9329	0.9914	0.9860	0.9086	0.9347	0.9967	0.9809	1				
39	S	0.9701	0.9772	0.9765	0.9746	0.9889	0.9636	0.7813	0.9793	0.9755	0.9946	0.9566	0.9823	0.9761	0.9835	1			
16	T	0.9689	0.9909	0.9848	0.9807	0.9887	0.9188	0.8919	0.9888	0.9773	0.9527	0.9505	0.9460	0.9854	0.9542	0.9920	1		
18	V	0.9769	0.9789	0.9799	0.9790	0.9950	1.0000	0.9678	0.9856	0.9764	0.8104	0.9117	0.9410	0.9691	0.9737	0.9661	0.9755	1	
5	Y	0.9194	0.7703	0.9351	0.8676	0.8922	0.5826	0.9508	0.8089	0.7482	0.8532	0.8544	0.8039	0.8180	0.8805	0.6635	0.8577	0.7915	1
116	X	0.9976	0.9831	0.9939	0.9906	0.9993	0.9102	0.8950	0.9897	0.9890	0.9355	0.9662	0.9960	0.9833	0.9987	0.9898	0.9702	0.9734	0.8190

## RESULTS: RANK CORRELATION WITHIN

Specie_αchain	Deleted aa	$r_{\min}$ (where)	$r_{\max}$ (where)	$r < 0.5$	$r \geq 0.5$	$r \geq 0.75$	$r \geq 0.95$	$r \geq 0.99$
BTα1	W	0.4905 (V-Y)	0.9987 (L-R)	2	151	140	77	9
BTα2	C, W	0.6857 (H-L)	0.9989 (G-R)	0	136	133	103	23
CLα1	W	0.5438 (C-Y)	0.9974 (K-R)	0	153	145	90	12
CLα2	W	0.5255 (L-Y)	0.9988 (G-R)	0	153	138	105	17
DRα1	W, Y	0.5959 (H-L)	0.9968 (E-R)	0	153	151	97	11
DRα2	C, W	0.5852 (H-L)	0.9978 (G-P)	0	153	145	100	24
HSα1	H, M, W, Y	0.7363 (C-L)	0.9983 (G-P)	0	120	119	72	9
HSα2	C, W	0.5033 (M-Y)	0.9989 (G-R)	0	153	136	98	21
RNα1	C, I, M, W, Y	0.8953 (A-N)	0.9993 (G-X)	0	105	105	95	25
RNα2	C, M, W, Y	0.8709 (S-T)	0.9983 (G-P)	0	120	120	100	19

BTα1 = *Bos taurus* TICα1; BTα2 = *Bos taurus* TICα2; TIC = type I collagen;

aa = amino acid (one-letter abbreviation)

CLα1 = *Canis lupus* TICα1; CLα2 = *Canis lupus* TICα2;

DRα1 = *Danio rerio* TICα1; DRα2 = *Danio rerio* TICα2;

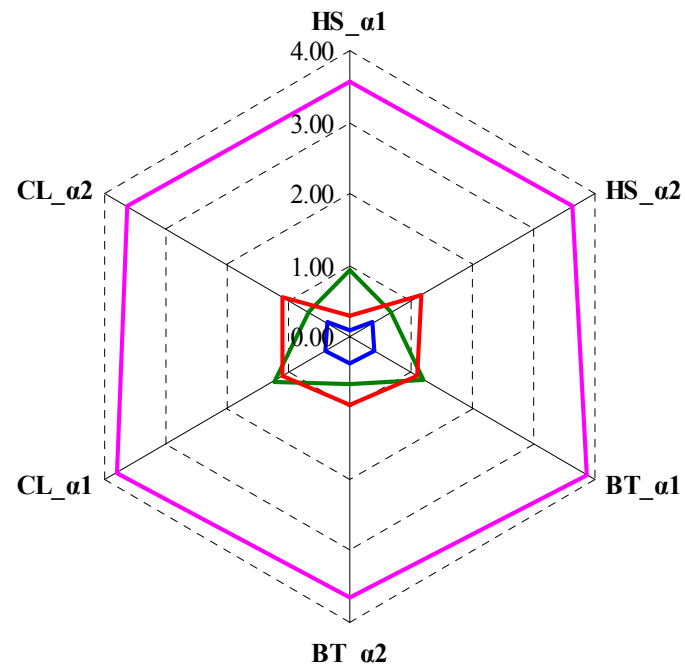
HSα1 = *Homo sapiens* TICα1; HSα2 = *Homo sapiens* TICα2;

RNα1 = *Rattus norvegicus* TICα1; RNα2 = *Rattus norvegicus* TICα2.

# RESULTS: RANK CORRELATION BETWEEN

Perfect correlation (1.0000)

BT $\alpha$ 1C- HS $\alpha$ 1C  
 BT $\alpha$ 1W- HS $\alpha$ 1W  
 BT $\alpha$ 1Y- HS $\alpha$ 1Y  
 CL $\alpha$ 2C- HS $\alpha$ 2C  
 CL $\alpha$ 2K- HS $\alpha$ 2K  
 CL $\alpha$ 2W- HS $\alpha$ 2W

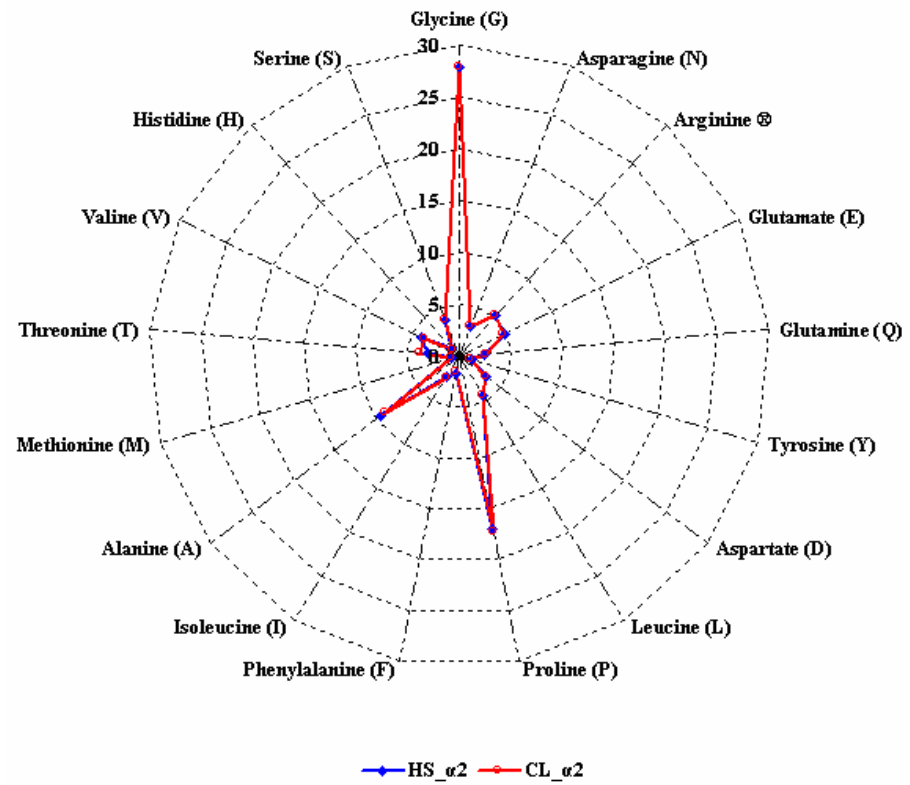


— Cysteine (C) — Tryptophan (W) — Tyrosine (Y) — Lysine (K)

# RESULTS: RANK CORRELATION BETWEEN

Good correlations ( $> 0.75$ )

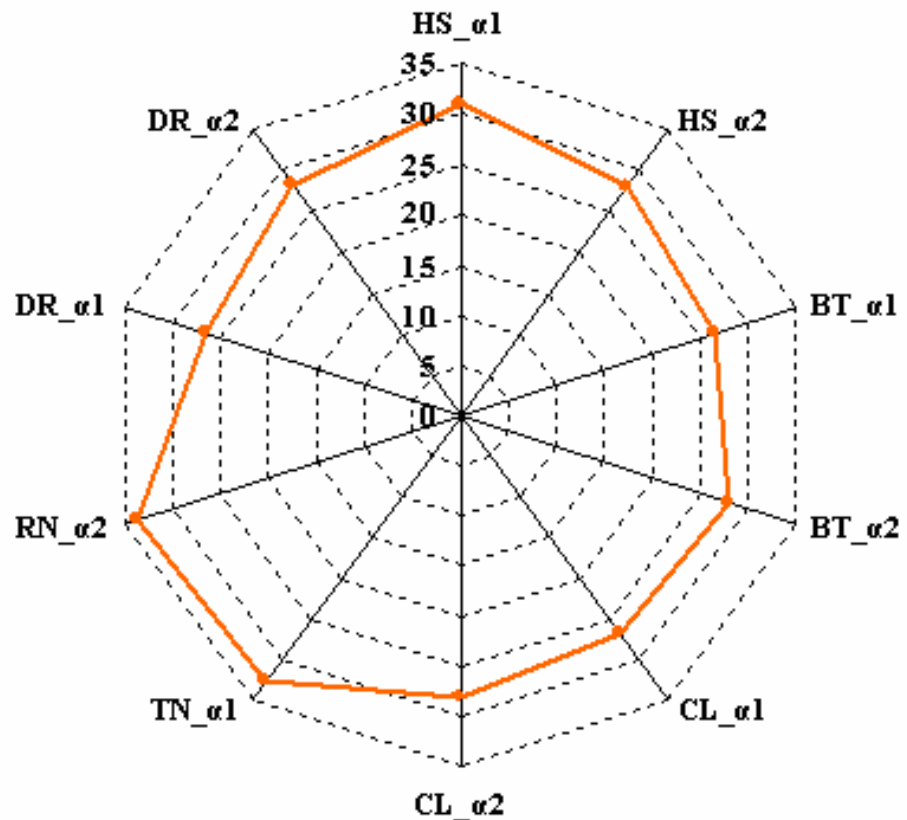
Amino acid	CL $\alpha$ 2-HS $\alpha$ 2
Glycine	0.993
Asparagine	0.988
Arginine	0.978
Glutamate	0.976
Glutamine	0.970
Tyrosine	0.968
Aspartate	0.952
Leucine	0.948
Proline	0.938
Phenylalanine	0.929
Isoleucine	0.876
Alanine	0.874
Methionine	0.857
Threonine	0.837
Valine	0.833
Histidine	0.818
Serine	0.760



# RESULTS: RANK CORRELATION BETWEEN

Good correlations ( $> 0.75$ )

	Glycine (G)
Ba2-Ra1	0.877
Ba1-Ra2	0.875
Ha2-Ra2	0.871
Da1-Ra2	0.870
Ca2-Ra2	0.868
Ba1-Da1	0.860
Ca1-Da1	0.841
Ba2-Ha1	0.835
Ba1-Ca1	0.830
Da1-Ha2	0.823
Ba1-Ca2	0.820
Ba1-Ha2	0.820
Ca2-Da1	0.819
Ca1-Ra2	0.814
Ha1-Ra1	0.806
Ca1-Ca2	0.763
Ca1-Ha2	0.763





## RESULTS: RANK CORRELATION BETWEEN

Weak correlations ( $0.50 < r < 0.75$ )

- o Histidine:
  - o BT $\alpha$ 1-HS $\alpha$ 1
  - o 0.666
  - o BT 0.62% - HS 0.28%
- o Thyrosine – Histidine
  - o BT $\alpha$ 2-RN $\alpha$ 1
  - o 0.577
  - o BT 0.95% - RN 0.28%

# RESULTS: AUTOCORRELATION

---

- Fifty-six out of one-hundred positive autocorrelations were identified
- The maximum number of amino acids identified on autocorrelation analysis:
  - alanine - type I collagen  $\alpha 1$  chain for *Canis lupus*
  - 9 out of 20 45% - 95%<sub>-fa</sub> CI [5 - 14].

# RESULTS: AUTOCORRELATION

---

- $\alpha 1$  type I collagen of *Bos Taurus* and  $\alpha 2$  type I collagen of *Homo sapiens*
  - 8 out of 20
  - 40% 95%<sub>-fa</sub> CI [5 - 14]
- The lowest performances:
  - type I collagen chains for *Ratus norvegicus*
  - One positive autocorrelation
  - Alanine for  $\alpha 1$  chain
  - Glycine for  $\alpha 2$  chain

# RESULTS: AUTOCORRELATION

- The dimension of the type I collagen substructures that autocorrelated:
  - 7 amino acids (*Ratus norvegicus*  $\alpha 2$  chain of type I collagen - glycine;  $r = 0.7300$ )
  - 1462 amino acids (*Bos Taurus*  $\alpha 1$  chain of type I collagen - leucine;  $r = 0.012$ ).
- The number of simultaneously presence of amino acid of interest in the same position on both substructures varied from 2 to 18
  - 0.012 - *Danio rerio* on 1037 amino acids

## RESULTS: AUTOCORRELATION

Chn	Siz	Smi	Sma	Spr	r	Chn	Siz	Smi	Sma	Spr	r
BTα1_A	380	26	27	3	0.0470	DRα1_A	314	26	27	5	0.1140
BTα1_D	271	16	16	2	0.0700	DRα1_D	214	17	17	3	0.1050
BTα1_E	28	3	4	2	0.5190	DRα1_G	161	18	18	3	0.0620
BTα1_L	43	8	8	5	0.5390	DRα1_I	1422	33	34	2	0.0370
BTα1_P	152	25	26	6	0.0810	DRα1_K	1405	55	55	4	0.0350
BTα1_Q	1449	50	50	2	0.0060	DRα1_L	12	4	5	3	0.4780
BTα1_T	1232	26	27	2	0.0550	DRα1_T	1414	51	52	2	0.0030
BTα1_V	1222	29	30	2	0.0450	DRα2_A	216	19	19	3	0.0770
BTα2_A	333	31	32	3	0.0010	DRα2_K	1310	47	47	3	0.0290
BTα2_K	1317	45	46	3	0.0330	DRα2_L	12	4	5	3	0.4780
BTα2_L	12	4	5	3	0.4780	DRα2_N	517	14	15	2	0.1130
BTα2_N	1141	27	28	2	0.0500	DRα2_P	74	12	12	2	0.0050
BTα2_P	49	5	6	2	0.2850	DRα2_S	1205	56	57	3	0.0070
BTα2_V	713	21	21	2	0.0680	HSα1_A	381	27	28	3	0.0400
CLα1_A	326	22	23	2	0.0210	HSα1_D	272	16	16	2	0.0700
CLα1_D	268	15	15	2	0.0820	HSα1_E	465	26	27	2	0.0200
CLα1_E	28	3	4	2	0.5190	HSα1_L	43	8	8	5	0.5390
CLα1_L	39	8	8	5	0.5280	HSα1_P	176	27	27	6	0.0810
CLα1_P	83	5	6	2	0.3200	HSα2_A	104	6	7	2	0.2630
CLα1_Q	1375	47	48	2	0.0080	HSα2_E	377	17	18	2	0.0710
CLα1_T	1229	29	30	2	0.0450	HSα2_K	1319	45	46	3	0.0330
CLα1_V	1219	29	30	2	0.0450	HSα2_L	12	4	5	3	0.4780
CLα1_Y	1211	5	6	2	0.3620	HSα2_N	1143	26	27	2	0.0540
CLα2_A	335	31	32	3	0.0010	HSα2_P	49	5	6	2	0.2850
CLα2_K	1319	45	46	3	0.0330	HSα2_S	1219	43	44	2	0.0110
CLα2_L	12	4	5	3	0.4780	HSα2_V	752	27	28	3	0.0750
CLα2_N	1143	26	27	2	0.0540	RNα1_A	184	16	16	2	0.0420
CLα2_P	49	5	6	2	0.2850	RNα2_G	7	2	3	2	0.7300

Chn = the abbreviation of the species, type I collagen chain (α1/α2), amino acid (one letter abbreviation, see MATERIAL AND METHODS - type I collagen);  
 BTα1<sub>i</sub> = *Bos taurus* TICα1; BTα2<sub>i</sub> = *Bos taurus* TICα2; CLα1<sub>i</sub> = *Canis lupus* TICα1; CLα2<sub>i</sub> = *Canis lupus* TICα2; DRα1<sub>i</sub> = *Danio rerio* TICα1;  
 DRα2<sub>i</sub> = *Danio rerio* TICα2; HSα1<sub>i</sub> = *Homo sapiens* TICα1; HSα2<sub>i</sub> = *Homo sapiens* TICα2; RNα1<sub>i</sub> = *Rattus norvegicus* TICα1;  
 RNα2<sub>i</sub> = *Rattus norvegicus* TICα2; i = one letter abbreviation of standard amino acids; TIC = type I collagen;  
 Siz = the dimension of the collagen type I substructures (number of amino acids) that autocorrelated;  
 Smi and Sma = number of amino acids present in the two substructures (one being higher than other);  
 Spr = number of simultaneously presence of amino acid of interest in both substructures (i.e. the same position);  
 r = correlation coefficient.



# RESULTS: AUTOCORRELATION

---

- o  $r = 0.5390$ :
  - o *Bos taurus* and *Homo sapiens*  $\alpha 1$  chain of type I collagen - leucine.
  - o Dimension: 43 aa, 5 leucine simultaneously in the same position.
- o  $r = 0.5280$ :
  - o *Canis lupus*  $\alpha 1$  chain of type I collagen – leucine
  - o Dimension: 39 aa, 5 leucine simultaneously in the same position.
- o  $r = 0.5190$ :
  - o *Bos taurus* and *Canis lupus*  $\alpha 1$  chain of type I collagen - glutamate
  - o Dimension: 28, 2 glutamate simultaneously in the same position

# CONCLUDING REMARKS


---

- o The rank correlation analysis revealed the existence of a moderate to a very good correlation between ranks of standard amino acids position in the investigated type I collagen chains on all species.
- o The autocorrelation is not related with the frequency distribution of amino acids.

# ACKNOWLEDGEMENTS

- o FICAMC'08 organizers
- o Grant: RO/CNCSIS/AT/93
  - o [Sorana D. Bolboacă](#) (principal investigator)
  - o [Lorentz Jäntschi](#) (co-investigator)





*A statistical analysis, properly conducted, is a delicate dissection of uncertainties, a surgery of suppositions.*

**M. J. Moroney**

**THANK YOU FOR ATTENTION!**