

Table Ts1. Fraction of all 400 amino acid dimers in the human genomic database containing 37391 sequences with 17,863,015 amino acid residues as total. Please see at the end of this file for a full output from the program used for the production of the analysis shown below.

Analysis of the human genomic database

Diad Total Fraction

```
-----
```

1	AA	134116	0.007498345
2	AR	72892	0.004075348
3	AN	34598	0.001934353
4	AD	51028	0.002852945
5	AC	25965	0.001451688
6	AQ	60134	0.003362056
7	AE	86126	0.004815253
8	AG	95097	0.005316816
9	AH	29752	0.001663417
10	AI	48081	0.002688180
11	AL	130653	0.007304731
12	AK	58899	0.003293008
13	AM	24374	0.001362736
14	AF	48445	0.002708531
15	AP	84943	0.004749112
16	AS	106658	0.005963185
17	AT	66639	0.003725746
18	AW	15826	0.000884822
19	AY	26175	0.001463428
20	AV	85514	0.004781037
21	RA	77148	0.004313299
22	RR	86555	0.004839238
23	RN	37026	0.002070101
24	RD	49091	0.002744648
25	RC	23349	0.001305429
26	RQ	47744	0.002669338
27	RE	72588	0.004058352
28	RG	74481	0.004164188
29	RH	29703	0.001660677
30	RI	42871	0.002396892
31	RL	99137	0.005542690
32	RK	64875	0.003627122
33	RM	19024	0.001063620
34	RF	32431	0.001813198
35	RP	65675	0.003671850
36	RS	79568	0.004448599
37	RT	51026	0.002852833
38	RW	13695	0.000765679
39	RY	24519	0.001370843
40	RV	59081	0.003303183
41	NA	35339	0.001975782
42	NR	32004	0.001789324

43	NN	25478	0.001424460
44	ND	25417	0.001421049
45	NC	14747	0.000824496
46	NQ	27976	0.001564121
47	NE	38387	0.002146194
48	NG	42154	0.002356805
49	NH	16684	0.000932792
50	NI	35551	0.001987635
51	NL	63827	0.003568529
52	NK	37124	0.002075581
53	NM	14155	0.000791398
54	NF	24959	0.001395443
55	NP	37851	0.002116227
56	NS	53839	0.003010106
57	NT	33727	0.001885656
58	NW	7872	0.000440119
59	NY	19196	0.001073237
60	NV	37082	0.002073232
61	DA	51416	0.002874638
62	DR	42536	0.002378162
63	DN	28132	0.001572843
64	DD	44058	0.002463256
65	DC	18601	0.001039971
66	DQ	31301	0.001750020
67	DE	57752	0.003228880
68	DG	58057	0.003245932
69	DH	19891	0.001112094
70	DI	43039	0.002406285
71	DL	85344	0.004771532
72	DK	42016	0.002349089
73	DM	17505	0.000978694
74	DF	35757	0.001999152
75	DP	51689	0.002889901
76	DS	72735	0.004066570
77	DT	42963	0.002402036
78	DW	10880	0.000608294
79	DY	24136	0.001349429
80	DV	53062	0.002966664
81	CA	23394	0.001307945
82	CR	25154	0.001406345
83	CN	14598	0.000816165
84	CD	17769	0.000993454
85	CC	13528	0.000756342
86	CQ	20242	0.001131718
87	CE	23776	0.001329302
88	CG	36130	0.002020007
89	CH	12527	0.000700377
90	CI	16680	0.000932569
91	CL	39618	0.002215018
92	CK	21327	0.001192380
93	CM	6528	0.000364977
94	CF	15184	0.000848928

95	CP	27010	0.001510113
96	CS	37514	0.002097385
97	CT	19976	0.001116846
98	CW	5761	0.000322094
99	CY	10325	0.000577265
100	CV	23752	0.001327960
101	QA	64853	0.003625892
102	QR	55485	0.003102133
103	QN	31828	0.001779484
104	QD	40752	0.002278420
105	QC	17593	0.000983614
106	QQ	54172	0.003028724
107	QE	70352	0.003933338
108	QG	54616	0.003053548
109	QH	23961	0.001339645
110	QI	33717	0.001885097
111	QL	84009	0.004696893
112	QK	51215	0.002863400
113	QM	17328	0.000968798
114	QF	23347	0.001305317
115	QP	53172	0.002972815
116	QS	55321	0.003092964
117	QT	39905	0.002231065
118	QW	9829	0.000549534
119	QY	19061	0.001065689
120	QV	50335	0.002814200
121	EA	95182	0.005321568
122	ER	71967	0.004023632
123	EN	53947	0.003016145
124	ED	76689	0.004287636
125	EC	28440	0.001590063
126	EQ	54671	0.003056623
127	EE	139892	0.007821278
128	EG	76844	0.004296302
129	EH	26378	0.001474778
130	EI	53110	0.002969348
131	EL	111496	0.006233674
132	EK	95508	0.005339795
133	EM	27245	0.001523252
134	EF	34391	0.001922780
135	EP	59728	0.003339357
136	ES	76006	0.004249450
137	ET	60015	0.003355403
138	EW	12552	0.000701775
139	EY	26877	0.001502677
140	EV	73413	0.004104477
141	GA	90157	0.005040624
142	GR	74742	0.004178781
143	GN	41676	0.002330080
144	GD	56529	0.003160503
145	GC	25399	0.001420043
146	GQ	52353	0.002927025

147	GE	79357	0.004436802
148	GG	98839	0.005526029
149	GH	32079	0.001793518
150	GI	48317	0.002701374
151	GL	110989	0.006205328
152	GK	70387	0.003935295
153	GM	21281	0.001189808
154	GF	42785	0.002392084
155	GP	85896	0.004802394
156	GS	109231	0.006107040
157	GT	66284	0.003705899
158	GW	15573	0.000870677
159	GY	30059	0.001680581
160	GV	65792	0.003678391
161	HA	24731	0.001382695
162	HR	30419	0.001700708
163	HN	15039	0.000840821
164	HD	15519	0.000867658
165	HC	13471	0.000753156
166	HQ	25776	0.001441121
167	HE	23415	0.001309119
168	HG	29703	0.001660677
169	HH	17003	0.000950628
170	HI	21686	0.001212451
171	HL	53730	0.003004012
172	HK	23128	0.001293073
173	HM	9536	0.000533152
174	HF	19159	0.001071168
175	HP	31895	0.001783230
176	HS	42520	0.002377268
177	HT	30575	0.001709430
178	HW	6557	0.000366598
179	HY	13836	0.000773562
180	HV	26796	0.001498148
181	IA	44101	0.002465660
182	IR	40854	0.002284123
183	IN	30470	0.001703559
184	ID	32965	0.001843053
185	IC	19831	0.001108739
186	IQ	39117	0.002187008
187	IE	42725	0.002388729
188	IG	37352	0.002088328
189	IH	26494	0.001481264
190	II	40348	0.002255833
191	IL	78755	0.004403145
192	IK	43626	0.002439104
193	IM	15467	0.000864751
194	IF	32537	0.001819124
195	IP	45681	0.002553997
196	IS	63855	0.003570095
197	IT	44666	0.002497249
198	IW	9040	0.000505421

199	IY	24072	0.001345851
200	IV	42578	0.002380510
201	LA	124022	0.006933996
202	LR	109157	0.006102902
203	LN	59428	0.003322584
204	LD	81569	0.004560474
205	LC	40205	0.002247837
206	LQ	102974	0.005757215
207	LE	124896	0.006982861
208	LG	110942	0.006202701
209	LH	49392	0.002761477
210	LI	66334	0.003708694
211	LL	196321	0.010976190
212	LK	98253	0.005493266
213	LM	31111	0.001739397
214	LF	59003	0.003298822
215	LP	112491	0.006289304
216	LS	141489	0.007910565
217	LT	90747	0.005073610
218	LW	21376	0.001195119
219	LY	43237	0.002417355
220	LV	98606	0.005513002
221	KA	71962	0.004023352
222	KR	57748	0.003228656
223	KN	40094	0.002241632
224	KD	53167	0.002972535
225	KC	22285	0.001245941
226	KQ	44190	0.002470636
227	KE	87397	0.004886314
228	KG	58421	0.003266283
229	KH	24679	0.001379788
230	KI	47275	0.002643117
231	KL	86970	0.004862440
232	KK	80068	0.004476554
233	KM	22960	0.001283680
234	KF	29093	0.001626572
235	KP	57946	0.003239726
236	KS	67304	0.003762926
237	KT	53634	0.002998645
238	KW	10791	0.000603318
239	KY	26116	0.001460130
240	KV	59244	0.003312296
241	MA	35068	0.001960631
242	MR	18996	0.001062055
243	MN	15348	0.000858097
244	MD	21368	0.001194672
245	MC	7364	0.000411717
246	MQ	16270	0.000909646
247	ME	32156	0.001797823
248	MG	23712	0.001325724
249	MH	8276	0.000462706
250	MI	14166	0.000792013

251	ML	34575	0.001933067
252	MK	24845	0.001389069
253	MM	9059	0.000506483
254	MF	12332	0.000689475
255	MP	20010	0.001118747
256	MS	27841	0.001556574
257	MT	19751	0.001104267
258	MW	4509	0.000252095
259	MY	9975	0.000557696
260	MV	23894	0.001335899
261	FA	34061	0.001904330
262	FR	35660	0.001993729
263	FN	23095	0.001291228
264	FD	27657	0.001546286
265	FC	17057	0.000953647
266	FQ	32071	0.001793070
267	FE	33343	0.001864187
268	FG	38358	0.002144573
269	FH	18721	0.001046680
270	FI	31152	0.001741690
271	FL	71674	0.004007251
272	FK	30613	0.001711554
273	FM	11953	0.000668285
274	FF	27458	0.001535160
275	FP	37161	0.002077649
276	FS	61497	0.003438260
277	FT	35635	0.001992331
278	FW	9081	0.000507713
279	FY	20112	0.001124450
280	FV	36227	0.002025430
281	PA	97392	0.005445128
282	PR	69153	0.003866303
283	PN	31776	0.001776577
284	PD	49227	0.002752252
285	PC	22449	0.001255110
286	PQ	53714	0.003003117
287	PE	80747	0.004514516
288	PG	107570	0.006014174
289	PH	27610	0.001543659
290	PI	33408	0.001867821
291	PL	98395	0.005501206
292	PK	49355	0.002759408
293	PM	17051	0.000953311
294	PF	34853	0.001948610
295	PP	117339	0.006560353
296	PS	105989	0.005925782
297	PT	58084	0.003247441
298	PW	14096	0.000788099
299	PY	28715	0.001605438
300	PV	69088	0.003862669
301	SA	98026	0.005480575
302	SR	85459	0.004777961

303	SN	46276	0.002587263
304	SD	65251	0.003648144
305	SC	35332	0.001975391
306	SQ	69345	0.003877037
307	SE	91386	0.005109336
308	SG	103433	0.005782877
309	SH	39559	0.002211720
310	SI	55626	0.003110016
311	SL	146926	0.008214545
312	SK	69551	0.003888555
313	SM	25032	0.001399524
314	SF	54409	0.003041975
315	SP	111805	0.006250951
316	SS	168982	0.009447683
317	ST	81479	0.004555442
318	SW	21020	0.001175216
319	SY	35995	0.002012459
320	SV	85015	0.004753137
321	TA	69971	0.003912037
322	TR	45903	0.002566409
323	TN	29661	0.001658329
324	TD	41939	0.002344784
325	TC	24019	0.001342888
326	TQ	40944	0.002289155
327	TE	61339	0.003429427
328	TG	68168	0.003811232
329	TH	25496	0.001425466
330	TI	41059	0.002295584
331	TL	93598	0.005233008
332	TK	45589	0.002548854
333	TM	17848	0.000997871
334	TF	37218	0.002080836
335	TP	67374	0.003766840
336	TS	82572	0.004616551
337	TT	58778	0.003286243
338	TW	13013	0.000727549
339	TY	24874	0.001390690
340	TV	67307	0.003763094
341	WA	15885	0.000888121
342	WR	14757	0.000825055
343	WN	9829	0.000549534
344	WD	11462	0.000640834
345	WC	4530	0.000253270
346	WQ	9801	0.000547968
347	WE	15208	0.000850270
348	WG	14803	0.000827627
349	WH	5556	0.000310633
350	WI	9484	0.000530245
351	WL	22543	0.001260366
352	WK	14662	0.000819744
353	WM	5215	0.000291568
354	WF	7886	0.000440902

355	WP	10981	0.000613941
356	WS	16313	0.000912050
357	WT	12166	0.000680194
358	WW	3681	0.000205802
359	WY	5898	0.000329754
360	WV	13129	0.000734035
361	YA	23605	0.001319741
362	YR	28036	0.001567476
363	YN	18280	0.001022024
364	YD	20994	0.001173762
365	YC	11501	0.000643014
366	YQ	21649	0.001210383
367	YE	29026	0.001622826
368	YG	28535	0.001595375
369	YH	12893	0.000720840
370	YI	24182	0.001352001
371	YL	45601	0.002549525
372	YK	26242	0.001467175
373	YM	9426	0.000527002
374	YF	20808	0.001163363
375	YP	22942	0.001282673
376	YS	37560	0.002099957
377	YT	25585	0.001430442
378	YW	6354	0.000355248
379	YY	15582	0.000871180
380	YV	27120	0.001516263
381	VA	78053	0.004363896
382	VR	54656	0.003055784
383	VN	38053	0.002127521
384	VD	50102	0.002801173
385	VC	26274	0.001468964
386	VQ	48318	0.002701430
387	VE	66960	0.003743693
388	VG	63059	0.003525591
389	VH	28950	0.001618577
390	VI	49794	0.002783953
391	VL	110675	0.006187773
392	VK	56665	0.003168106
393	VM	20761	0.001160735
394	VF	41838	0.002339137
395	VP	66789	0.003734133
396	VS	86550	0.004838958
397	VT	67009	0.003746433
398	VW	12923	0.000722517
399	VY	28072	0.001569489
400	VV	70793	0.003957994

Output from the Cube program (Macintosh computers only) available upon request
jalat@dsvidf.cea.fr

Triad_DataB > * * * * Genomic Database Analysis * * * *

Triad_DataB> Analysis of human database
Database used ::: Hsapa2.gbk
Sliding frame = 19

Min-Mina 43 5158
Max-Maxa 26267 6316
Max-Tot 17886079

Total of fractions 0.9999996000
Max-fraction 0.0014685721 SSS
Min-fraction 0.0000024041 MWW Ratio 610.86

Total-of-dimers-fractions 1.000
Max-dimer-fraction 0.0109761900 LL
Min-dimer-fraction 0.0002058025 WW Ratio 53.33

Overall number of entries 37391
Minimal length of protein 25 NP_0010303
Maximal length of protein 33423 NP_596869
Number of AA residues 17863015
Total number of lines 3993454
Sliding frame (AA) 19
Signal_peptides 2177
Transit_peptides 205
Mature_proteins 2699
Database contains::: 37609

General statistics of the database

>10000	10K-20K	20K-30K	30K-40K	40K-50K
2249	7169	5478	4953	3906
50K-60K	60K-70K	70K-100K	100K-150K	> 150K
3337	2278	4026	2561	1652
< 3pI	3-4pI	4-5pI	5-6pI	6-7pI
9	191	3580	6908	5919
7-8pI	8-9pI	9-10pI	10-11pI	>pI 11
4730	5729	4848	2711	2984

Distribution of hydrophobicity indexes

0-10%	10-20%	20-30%	30-40%	40-50%
5447	8163	9361	6873	3583
50-60%	60-70%	70-80%	80-90%	90-100%
1902	1167	687	408	18

Total of masses 37609 Total of pIs 37609 Total of HIs37609

A	R	N	D	C	Q	E	G	H	I
1280167	1047713	624395	831947	410844	851028	1254946	1214190	474957	754842
7.0834	5.7972	3.4549	4.6033	2.2733	4.7089	6.9438	6.7183	2.6280	4.1767
L	K	M	F	P	S	T	W	Y	V
1751498	1003711	377131	632056	1164098	1489662	956200	222862	456476	1064292
9.6914	5.5537	2.0867	3.4973	6.4412	8.2426	5.2908	1.2331	2.5258	5.8889

Group No. 1 5447

A R N D C Q E G H I

6.8834	8.1570	3.2521	4.4068	2.2980	5.4456	8.4882	7.5969	3.0314	2.8693
L	K	M	F	P	S	T	W	Y	V
7.1972	7.3086	2.1266	2.6406	7.3845	8.6482	4.9037	1.1832	2.1704	4.0059

Group No. 2 8163

A	R	N	D	C	Q	E	G	H	I
7.3602	7.0048	3.3396	4.6091	2.2937	4.9630	7.3474	7.3460	2.6600	3.4648
L	K	M	F	P	S	T	W	Y	V
8.6448	5.9908	2.1360	3.0251	7.4244	8.5216	5.1512	1.3134	2.3098	5.0899

Group No. 3 9361

A	R	N	D	C	Q	E	G	H	I
7.4939	6.3389	3.3572	4.5887	2.4570	4.7046	6.5969	7.2047	2.7122	3.9519
L	K	M	F	P	S	T	W	Y	V
9.6615	5.3544	2.1638	3.4603	6.8262	8.1453	5.1793	1.4009	2.5145	5.8838

Group No. 4 6873

A	R	N	D	C	Q	E	G	H	I
7.6699	5.8612	3.2269	4.4372	2.5617	4.3483	5.9945	7.1335	2.6469	4.3202
L	K	M	F	P	S	T	W	Y	V
10.6919	4.8524	2.2515	3.8082	6.5350	7.9115	5.2176	1.4822	2.5433	6.4997

Group No. 5 3583

A	R	N	D	C	Q	E	G	H	I
8.1041	5.4784	3.0777	4.0279	2.7311	4.1392	5.4382	7.1757	2.6810	4.5914
L	K	M	F	P	S	T	W	Y	V
11.6340	4.3736	2.3755	4.1230	6.1164	7.7150	5.2212	1.4888	2.4928	7.0130

Group No. 6 1902

A	R	N	D	C	Q	E	G	H	I
8.0821	5.0063	2.8840	3.5544	3.2088	3.8706	4.6653	7.3119	2.5669	5.0234
L	K	M	F	P	S	T	W	Y	V
12.4951	3.8275	2.4221	4.5063	5.6590	7.8444	5.2135	1.6530	2.6082	7.5961

Group No. 7 1167

A	R	N	D	C	Q	E	G	H	I
8.4323	4.3775	2.6923	2.8977	3.6033	3.3776	3.5713	7.0780	2.3699	5.7687
L	K	M	F	P	S	T	W	Y	V
13.0783	3.2668	2.6735	5.3929	5.1988	7.9471	5.4512	1.7850	3.1889	7.8475

Group No. 8 687

A	R	N	D	C	Q	E	G	H	I
8.3070	3.5742	2.6849	2.3856	3.0728	2.6929	2.9723	6.6437	2.4361	6.5581
L	K	M	F	P	S	T	W	Y	V
14.0662	2.9777	3.2474	6.1777	4.8981	7.8242	5.6591	1.6626	3.7258	8.4283

Group No. 9 408

A	R	N	D	C	Q	E	G	H	I
7.4816	3.1211	2.6586	2.2519	3.4719	2.3927	2.3137	5.7227	2.8160	7.9777
L	K	M	F	P	S	T	W	Y	V
14.1920	2.8001	3.5267	6.3269	4.5781	7.9442	6.0652	0.9667	4.2971	9.0869

Group No. 10 18

A	R	N	D	C	Q	E	G	H	I
---	---	---	---	---	---	---	---	---	---

7.9176	1.3882	1.0552	3.0372	4.2554	0.8711	4.8239	3.2153	3.7439	10.0689
L	K	M	F	P	S	T	W	Y	V
13.1135	1.6532	3.3890	2.5609	4.9687	5.3866	4.9072	0.5040	8.6948	14.4217

```

TAAr.f>      37391 Entries
              37391 Codes
              37391 Sequences
3711693 Info lines
3993454 Total

```

Distribution of pIs vs. mass

	0.	3.	4.	5.	6.	7.	8.	9.	10.	11.	14.
0- 10	1	57	244	178	133	270	266	270	358	472	
10- 20	0	75	727	809	582	722	891	1054	985	1324	
20- 30	0	25	579	840	692	611	784	750	550	647	
30- 40	0	11	419	784	638	678	1025	764	330	304	
40- 50	5	3	372	740	636	496	740	588	196	130	
50- 60	1	2	266	705	659	522	595	441	103	43	
60- 70	0	0	170	462	465	329	436	333	62	21	
70- 100	1	0	453	974	928	594	577	392	77	30	
100- 150	0	6	209	785	710	322	285	197	39	8	
150- 0	0	0	0	0	0	0	0	0	0	0	

Distribution of HI vs. mass

	0.	10.	20.	30.	40.	50.	60.	70.	80.	90.	100.
0- 10	815	347	324	326	201	120	82	28	6	0	
10- 20	1423	1665	1566	1085	632	384	241	127	46	0	
20- 30	739	1238	1353	979	533	304	197	83	48	4	
30- 40	460	953	1136	918	445	302	203	254	275	7	
40- 50	406	746	1042	795	405	232	186	73	16	5	
50- 60	333	658	939	646	379	172	102	94	13	1	
60- 70	336	473	625	443	230	75	70	25	1	0	
70- 100	514	888	1224	816	342	165	71	3	2	1	
100- 150	254	678	742	528	251	96	12	0	0	0	
150- 0	0	0	0	0	0	0	0	0	0	0	

-> Triad_DataB.f: (((Job finished)))