

Genomic Correlates of Hyperthermostability, an Update*

Received for publication, February 6, 2003
Published, JBC Papers in Press, February 24, 2003, DOI 10.1074/jbc.M301327200

Karsten Suhre and Jean-Michel Claverie‡

From the Structural and Genomic Information Laboratory, UPR 2589-CNRS, Institute of Structural Biology and Microbiology, CNRS, Marseille, France

It has been shown (Cambillau, C., and Claverie, J. M. (2000) *J. Biol. Chem.* 275, 32383–32386) that a large difference between the proportions of charged *versus* polar (non-charged) amino acids (CvP-bias) was an adequate, if empirical, signature of the proteome of hyperthermophilic organisms ($T_{\text{growth}} > 80^\circ\text{C}$). Since that study, the number of available microbial genomes has more than doubled, raising the possibility that the simple CvP-bias rule might no longer hold. Taking advantage of the new sequence data, we re-analyzed the genomes of 9 fully sequenced thermophiles, 9 hyperthermophiles, and 53 mesothermophile microorganisms to identify the genomic correlates of hyperthermostability on a wider data set. Our new results confirm that the CvP-bias previously identified on a much smaller data set still holds. Moreover, we show that it is an optimal criterion, in the sense that it corresponds to the most discriminating factor between hyperthermophilic and mesothermophilic microorganisms in a principal component analysis. In parallel, we evaluated two other recently proposed correlates of hyperthermostability, the proteome average pI and the dinucleotide statistical index (Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K., and Suzuki, M. (2000) *Proc. Natl. Acad. Sci.* 97, 14257–14262). We show that the CvP-bias is the sole criterion that is able to clearly discriminate hyperthermophile from mesothermophile microorganisms on a global genomic basis.

Although most organisms grow at temperatures ranging between 20 and 50 °C, several archaea and a few bacteria, such as *Pyrococcus* and *Aquifex*, have been found capable of withstanding temperatures close to or higher than 100 °C. Identification of the molecular basis of the increased thermostability of the proteins of such hyperthermophilic organisms is expected to help our understanding of protein folding as well as the design of enzymes retaining their activity at high temperatures (Ref. 1 and references therein). In a previous comparative study, Cambillau and Claverie (2) found that a large difference between the proportions of charged (Asp, Glu, Lys, Arg) *versus* polar (non-charged) (Asn, Gln, Ser, Thr) amino acids (abbreviated as CvP¹-bias) was the most promi-

nent signature of the hyperthermophilic life style at the proteome level. This global CvP-bias was reflected in the amino acid composition of the water-accessible residues computed from an analysis of the surface of 131 mesophilic *versus* 58 hyperthermophilic proteins.

Given the rapidly increasing number of fully sequenced microbial genomes (more than doubled since the initial study), inferences derived in the past from correlation studies on a limited set of examples are in constant danger of being proven wrong. Now the genomes of seven new thermophilic and hyperthermophile archaea and of three new thermophilic bacteria have been deciphered. Besides these thermophilic organisms, numerous new extremophile organisms, such as the halophilic archaea *Halobacterium* sp., the halophilic bacterium *Sinorhizobium meliloti*, and the alkaliphilic bacterium *Bacillus halodurans* have been added to the list, as well as a large number of mesophilic bacteria and archaea (Table I). With these new genomes from a wider evolutionary spectrum, any previously derived “life style” criterion is at risk of failure in the form of false positive (classifying a mesophile as hyperthermophile) or false negative (classifying an hyperthermophile as mesophile) predictions. In particular, the newly added extremophile sequence data allow us to investigate whether the bias in favor of charged rather than polar residues previously observed in the proteome of hyperthermophiles by Cambillau and Claverie (2) is truly specific to the adaptation in high temperatures or whether it could also be linked to other extreme environments.

This new body of sequence data also gives us an opportunity to reassess the conclusions of several other studies. By systematically comparing the archaeon *Thermoplasma volcanium* genomic sequence with seven other genomic sequences of archaea, all exhibiting higher optimal growth temperature (OGT), Kawashima *et al.* (3) identified a number of strong correlations between some characteristics of genome organization. For instance, they reported that the J_2 index (computed from the dinucleotide frequency, see “Materials and Methods”) was increasing together with the OGT. They also noticed some characteristic changes in the distribution of the isoelectric points (pI) of the proteins: with increasing OGT, the fraction of the basic protein subset (pI > 7) becomes larger and with it the genome-average pI.

In a different work, Kreil and Ouzounis (4) used hierarchical clustering and principal component analysis to identify the factors affecting the global amino acid composition of the predicted proteome of 6 thermophilic archaea, 2 thermophilic bacteria, 17 mesophilic bacteria, and 2 eukaryotic species. They concluded that the G + C content is indeed a dominant discriminating property but unrelated to any preference for a thermophilic lifestyle. They also noticed that thermophilic species could be identified by their global amino acid compositions alone, albeit without precisely defining the nature of this discriminating power.

In this study, we applied a similar approach to a much

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 33-4-91-16-45-48; Fax: 33-4-91-16-45-49; E-mail: Jean-Michel.Claverie@igs.cnrs-mrs.fr.

¹ The abbreviations used are: CvP, charged *versus* polar; OGT, optimal growth temperature; PCA, principal component analysis; G + C, guanine + cytosine.

TABLE I
Properties of the 71 genomes analyzed here, ranked by decreasing CvP bias

Properties of the 71 genomes analyzed here, ranked by decreasing CvP bias (see "Materials and Methods" for details). Mesophiles (OGT, <55%) are highlighted in blue, thermophiles (OGT, <80%) in orange, and hyperthermophiles in red.

genome	kingdom	OGT	G+C%	PCA 2	CvP-bias	J2	average pl
<i>Methanopyrus kandleri</i>	A	98°C	61,2	-6,07	17,99	0,06	5,94
<i>Pyrococcus abyssi</i>	A	97°C	44,7	-4,18	15,54	0,17	7,35
<i>Aquifex aeolicus</i>	B	90°C	43,3	-3,85	15,41	0,21	7,56
<i>Methanococcus jannaschii</i>	A	85°C	31,3	-2,91	14,85	0,13	7,30
<i>Pyrococcus horikoshii</i>	A	95°C	41,9	-3,84	14,67	0,17	7,67
<i>Pyrococcus furiosus</i>	A	98°C	40,8	-3,88	14,57	0,18	7,45
<i>Archaeoglobus fulgidus</i>	A	82°C	48,6	-3,78	14,05	0,15	6,62
<i>Thermotoga maritima</i>	B	80°C	46,3	-3,56	13,23	0,18	6,87
<i>Thermoanaerobacter tengcongensis</i>	B	75°C	37,6	-2,58	12,09	0,14	7,13
<i>Pyrobaculum aerophilum</i>	A	98°C	51,4	-3,01	11,99	0,06	8,02
<i>Methanobacterium thermoautotrophicum</i>	A	65°C	49,5	-2,88	11,05	0,10	6,02
<i>Aeropyrum pernix</i>	A	90°C	56,3	-3,26	10,62	0,11	8,09
<i>Clostridium perfringens</i>	B	37°C	28,5	-1,32	9,39	0,12	6,49
<i>Carboxydotherrmus hydrogenoformans</i>	B	68°C	42,0	-1,57	9,24	0,15	7,46
<i>Halobacterium sp</i>	A	37°C	65,9	-0,81	9,17	-0,07	4,68
<i>Fusobacterium nucleatum</i>	B	37°C	27,2	-0,79	9,06	0,13	7,20
<i>Bacillus stearothermophilus</i>	B	55°C	52,7	-0,45	9,05	0,01	8,25
<i>Sulfolobus tokodaii</i>	A	80°C	32,8	-1,60	8,90	0,10	7,66
<i>Sulfolobus solfataricus</i>	A	78°C	35,8	-1,62	8,81	0,10	7,93
<i>Streptomyces coelicolor</i>	B	28°C	72,0	-0,59	7,99	-0,04	6,68
<i>Methanosarcina mazei</i>	A	37°C	41,5	-1,26	7,91	0,14	6,39
<i>Thermoplasma volcanium</i>	A	60°C	39,9	-1,03	7,50	0,06	6,95
<i>Thermoplasma acidophilum</i>	A	58°C	46,0	-1,03	7,43	0,04	6,76
<i>Sinorhizobium meliloti</i>	B	26°C	62,2	-0,14	7,35	-0,01	6,93
<i>Caulobacter crescentus</i>	B	30°C	67,2	0,48	6,99	-0,02	7,35
<i>Pseudomonas aeruginosa</i>	B	37°C	66,6	0,72	6,66	-0,04	6,84
<i>Bacillus halodurans</i>	B	30°C	43,7	-0,58	6,62	0,07	6,43
<i>Brucella melitensis</i>	B	37°C	57,2	0,29	6,58	-0,02	7,28
<i>Methanosarcina acetivorans</i>	A	39°C	42,7	-0,90	6,56	0,14	6,57
<i>Chlorobium tepidum</i> TLS	B	48°C	56,5	-0,13	6,42	0,01	7,38
<i>Campylobacter jejuni</i>	B	37°C	30,6	1,43	6,30	0,13	7,44
<i>Clostridium acetobutylicum</i>	B	37°C	30,9	-0,11	6,28	0,08	7,36
<i>Mesorhizobium loti</i>	B	26°C	62,5	0,51	6,20	-0,04	7,27
<i>Listeria innocua</i>	B	37°C	37,4	-0,01	6,16	0,06	6,14
<i>Bacillus subtilis</i>	B	30°C	43,5	0,10	6,07	0,06	6,59
<i>Listeria monocytogenes</i>	B	37°C	38,0	-0,06	6,00	0,06	5,97
<i>Agrobacterium tumefaciens</i> C58	B	30°C	59,0	0,39	5,93	-0,01	6,95
<i>Streptococcus pneumoniae</i> R6	B	37°C	39,7	0,37	5,52	0,10	6,45
<i>Borrelia burgdorferi</i>	B	35°C	28,2	1,25	5,51	0,14	8,37
<i>Treponema pallidum</i>	B	37°C	52,8	0,28	4,79	-0,03	8,17
<i>Neisseria meningitidis</i> MC58	B	37°C	51,5	1,15	4,63	0,00	7,37
<i>Deinococcus radiodurans</i>	B	30°C	66,6	0,71	4,32	-0,03	7,12
<i>Ralstonia solanacearum</i>	B	30°C	67,0	1,37	4,27	-0,11	7,59
<i>Streptococcus pyogenes</i>	B	37°C	38,5	1,25	4,20	0,08	6,71
<i>Lactococcus lactis</i>	B	30°C	35,3	0,83	4,11	0,11	6,62
<i>Helicobacter pylori</i> 26695	B	37°C	38,9	1,87	4,05	0,13	7,98
<i>Mycoplasma pulmonis</i>	B	37°C	26,6	1,97	3,97	0,18	8,63
<i>Mycobacterium tuberculosis</i> H37Rv	B	37°C	65,6	0,56	3,93	-0,08	6,85
<i>Corynebacterium glutamicum</i>	B	30°C	53,8	0,47	3,91	0,01	5,62
<i>Mycobacterium leprae</i>	B	37°C	57,8	0,35	3,83	-0,07	6,67
<i>Xanthomonas campestris</i>	B	26°C	65,1	1,95	3,43	-0,13	7,28
<i>Staphylococcus aureus</i> Mu50	B	37°C	32,9	1,27	3,42	0,00	6,73
<i>Haemophilus influenzae</i>	B	37°C	38,2	1,56	3,12	0,02	6,98
<i>Salmonella typhi</i>	B	37°C	51,9	1,53	2,82	-0,04	6,97
<i>Escherichia coli</i> K12	B	37°C	50,8	1,54	2,63	-0,04	6,81
<i>Salmonella typhimurium</i> LT2	B	37°C	52,2	1,63	2,59	-0,04	6,94
<i>Rickettsia conorii</i>	B	37°C	32,4	1,78	2,51	0,04	7,84
<i>Pasteurella multocida</i>	B	37°C	40,4	1,96	2,43	0,01	7,02
<i>Buchnera sp</i>	B	26°C	26,4	1,96	2,31	0,08	9,44
<i>Rickettsia prowazekii</i>	B	37°C	29,0	1,85	2,19	0,02	8,26
<i>Vibrio cholerae</i>	B	28°C	47,5	1,90	1,91	-0,01	6,83
<i>Xylella fastidiosa</i>	B	26°C	52,6	1,76	1,73	-0,06	7,71
<i>Chlamydia trachomatis</i>	B	37°C	41,3	1,47	1,66	0,14	7,12
<i>Chlamydia muridarum</i>	B	37°C	40,3	1,52	1,51	0,15	7,29
<i>Ureaplasma urealyticum</i>	B	37°C	25,5	3,04	1,43	0,06	8,30
<i>Chlamydia pneumoniae</i> AR39	B	37°C	40,6	1,60	1,29	0,14	7,35
<i>Yersinia pestis</i>	B	37°C	47,6	2,11	1,07	-0,02	7,05
<i>Mycoplasma genitalium</i>	B	37°C	31,7	3,22	0,60	0,11	9,00
<i>Synechocystis</i> PCC6803	B	37°C	47,7	2,14	0,35	0,10	6,34
<i>Mycoplasma pneumoniae</i>	B	37°C	40,0	3,25	-0,21	0,07	8,57
<i>Nostoc sp</i>	B	30°C	41,3	2,41	-0,99	0,05	6,80

larger genomic data set. This allows us to show that the most discriminating factor between hyperthermophile and mesophile genomes remains the absolute difference between the frequency of charged and polar amino acid residues (CvP-bias). Further-

more, we show that a high J_2 index value is a necessary, but not sufficient, criterion for the recognition of a hyperthermophilic proteome, whereas the previously proposed genome-average pl is not a satisfactory correlate of hyperthermostability.

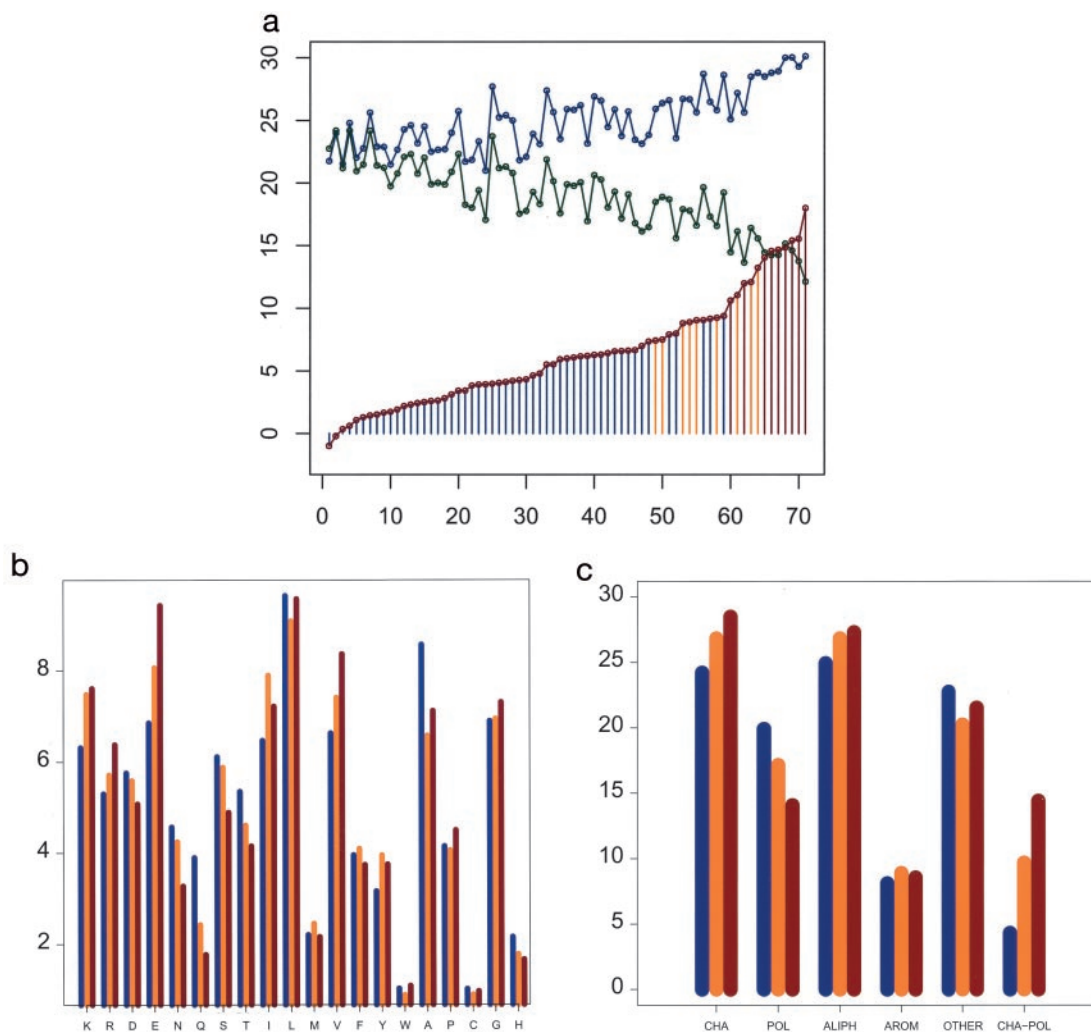


FIG. 1. *a*, plot of the percentages of charged amino acids (Asp, Glu, Lys, Arg, blue), polar non-charged amino acids (Asn, Gln, Ser, Thr, green), and of the difference of the two (CvP-bias, red). Blue, orange, and red vertical lines identify the mesophiles, thermophiles, and hyperthermophiles, respectively. *b*, plot of the percentages of the various amino acids in mesophiles (blue), thermophiles (orange), and hyperthermophiles (red). *c*, plot of the various amino acid classes; colors as in panel *b*.

MATERIALS AND METHODS

A total of 71 fully sequenced genomes was analyzed (Table I), including 53 mesophiles (<50 °C) (50 bacteria, 3 archaea), 9 thermophiles (50–80 °C) (4 bacteria, 5 archaea), and 9 hyperthermophiles (>80 °C) (1 bacteria, 8 archaea). 69 of those genome sequences were downloaded from GenBankTM and 2 unfinished genomes, *Carboxydotherrmus hydrogenoformans* and *Bacillus stearothermophilus*, from the Institute for Genome Research (TIGR, Rockville, MD) and from the University of Oklahoma, Norman, OK, respectively. The proteomes of the various organisms were defined according to the available open reading frame (ORF) annotation (when available in GenBankTM). For the two unfinished genomes, the proteomes were defined as the subsets of ORFs longer than 200 amino acids. Transmembrane segments were then predicted using the simple algorithm of Kyte and Doolittle (5). For each organism, the “soluble” moiety of the proteome was then defined by discarding all proteins (ORFs) containing at least two predicted transmembrane segments. The subsequent statistical analyses were all performed on these predicted soluble proteins.

Following Kawashima *et al.* (3), the genome sequences were used to compute the J_2 index as: $J_2 = F_{YY} + F_{RR} - F_{YR} - F_{RY}$ where F_{YY} designates the relative frequency of dinucleotides pyrimidine (TT, TC, CT, CC), F_{RR} the relative frequency of purine dinucleotides (AA, AG, GA, GG), and F_{YR} and F_{RY} the corresponding mixed purine/pyrimidine combinations. Pure purine and pyrimidine dinucleotides were found to be more frequent than their mixed counterparts in hyperthermophilic Pyrococci (3).

The program “iep” from the European Molecular Biology Open Software Suite (EMBOSS) was used to calculate the theoretical pI value for every soluble protein and, from those, the average pI for

every genome. Principal component analysis was performed using the statistical package *R*. 71 genomes from meso-, thermo-, and hyperthermophilic organisms were included. When multiple genome sequences from closely related strains or species were available, only one of them was included in the analysis to avoid introducing a potential statistical bias.

RESULTS

Hyperthermophiles Do Exhibit a Specific Proteome Composition Signature—The computed amino acid compositions for 9 fully sequenced hyperthermophilic and 53 mesophilic bacteria and archaea are presented in Table I and Fig. 1, *a–c*. The case of the 9 moderately thermophilic genomes will be discussed later. As previously observed by Cambillau and Claverie (Fig. 2 and *a–c* in Ref. 2) but now confirmed on a much larger data set, the proteins of hyperthermophiles exhibit a strong bias for the use of charged residues at the expense of polar residues. This includes strongly reduced frequencies for the thermolabile amino acid residues Gln and Asn, as also noted by Vieille *et al.* (6). Among other trends, the aliphatic residue Val appears to be preferred in hyperthermophiles, whereas the tiny non-polar non-charged residue Ala is avoided.

Principal component analysis (PCA) offers a way to analyze the data set more objectively and eventually identify more intricate relationships between amino acid frequencies and the adaptation of proteins to high temperature. Our PCA

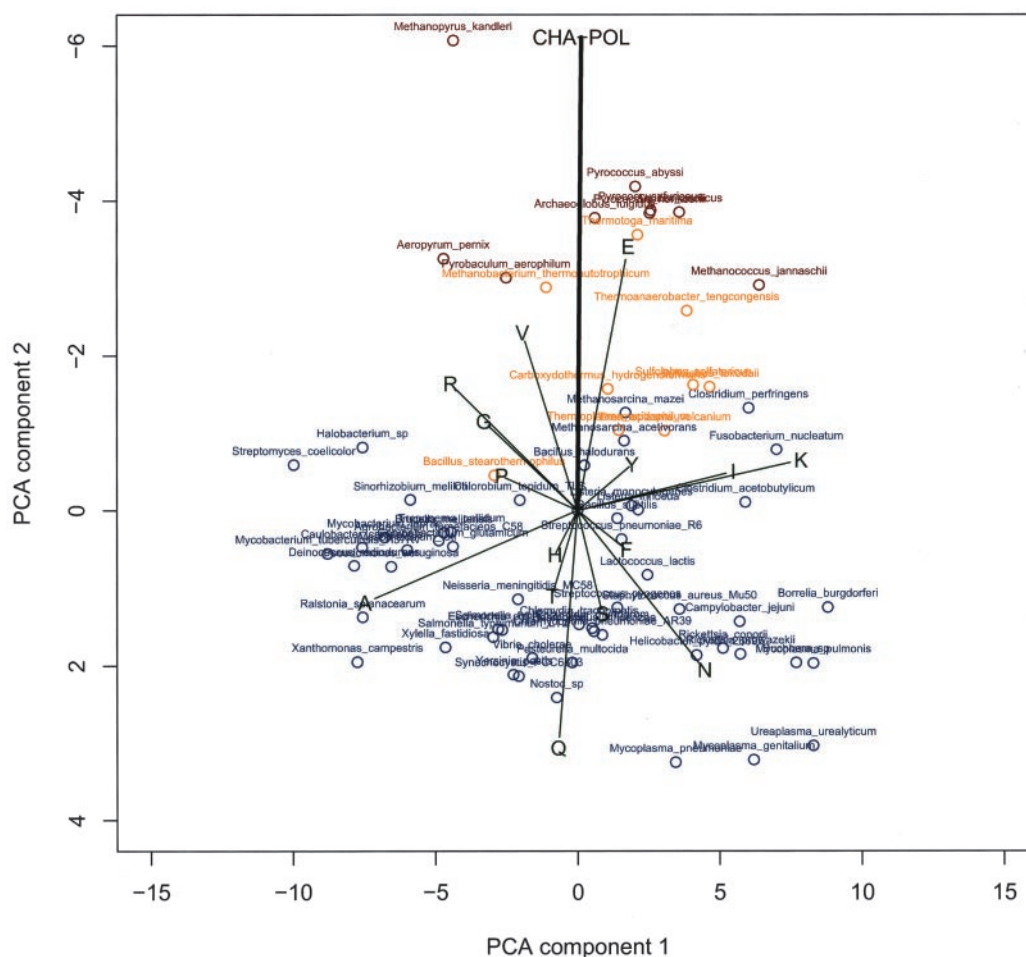


FIG. 2. **Biplot of a principal component analysis of the amino acid composition of all studied organisms.** *Blue dots* identify mesophiles; *orange dots* identify thermophiles; *red dots* identify hyperthermophiles. *Green vectors* represent the position of the different amino acids in the biplot (only vectors with significant contributions to PCA components 1 and 2 are drawn). The *black vector* represents $F_{CvP} = F_K + F_R + F_D + F_E - F_N - F_Q - F_S - F_T$.

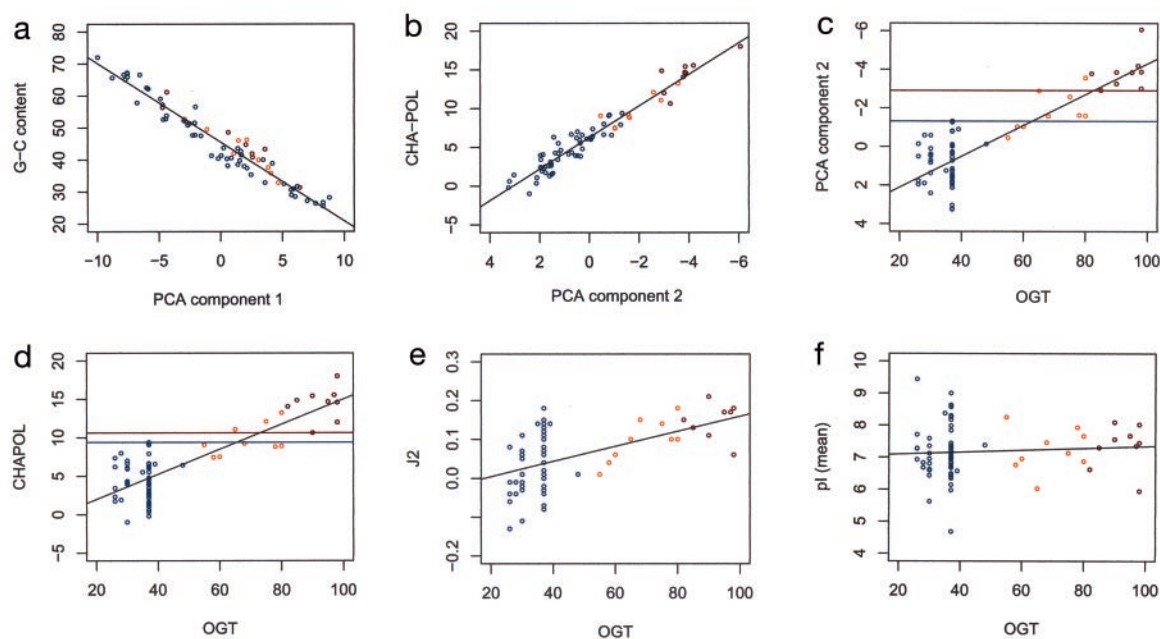


FIG. 3. **Correlation between various parameters.** *a*, correlation between G + C content and PCA component 1: $r^2 = 0.94$. *b*, correlation between *CuP*-bias and PCA component 2: $r^2 = 0.93$. *c*, correlation between PCA component 2 and OGT: $r^2 = 0.72$. *Red and blue lines* indicate the separation between the hyperthermophile organism with the lowest OGT and the mesophile with the highest OGT. *d*, correlation between *CuP*-bias and OGT: $r^2 = 0.68$. *e*, correlation between J_s and OGT: $r^2 = 0.27$. *f*, correlation between average proteomic pI and OGT: $r^2 = 0.01$.

analysis (Fig. 2) is consistent with earlier results obtained with a much smaller data set by Kreil and Ouzounis (4); 85% of the data variance is accounted by the first two principal components. The first component (72% of the variance) strongly correlates with the G + C content of the genomes (Fig. 3a, $r^2 = 0.94$). The second component (13% of the variance) turns out to be the sole discriminating factor between meso- and hyperthermophilic lifestyles, correlating quite well with the optimal growth temperature (Fig. 3c, OGT, $r^2 = 0.72$). All other PCA components account for not more than 3.7% each of the variance.

The first PCA component being the G + C content, we had to find the biochemical interpretation of the second component. First, we found that this component exhibits a strong correlation with the previously defined (CvP-bias) (Fig. 3b, $r^2 = 0.93$). Second, once projected on the first two dimensions of PCA space, the (20-dimensional) vector representing the difference between the frequencies of charged and polar residues (CHA-POL) appears almost parallel to the second PCA dimension (Fig. 2, black vector). As shown in Fig. 3d, the CvP-bias also correlates with the OGT almost as well as the second PCA component ($r^2 = 0.68$ and 0.72 , respectively). Finally, this parameter also successfully discriminates all hyperthermophilic microorganisms from all mesothermophilic ones. In conclusion, this analysis indicates that, among all possible combinations of G + C content and amino acid frequencies, the CvP-bias is a near optimal discriminating quantity to characterize the hyperthermophilic proteins from evolutionarily diverse microorganisms.

In contrast to the CvP-bias, other previously proposed quantities thought to characterize the hyperthermophilic life style do not fare well when confronted with this expanded proteome data set. Fig. 3, e and f, shows that the J_2 index, as well as the average pI, now fail to correctly discriminate between hyperthermophiles and mesophiles. In fact, both parameters correlate only weakly, if at all, with OGT ($r^2 = 0.27$ for J_2 and $r^2 = 0.01$ for average pI). However, all hyperthermophiles have a J_2 index greater than 0.06, so that a high J_2 index can be considered a necessary, but not a sufficient, criterion for the identification of a hyperthermophilic genome.

Moderately Thermophile Organisms—Because only a single moderate thermophilic organism had been sequenced (*Methanobacterium thermoautotrophicum*) at the time of the work of Cambillau and Claverie (2), little could be said about the properties of moderately thermophile organisms. Proteome data is now available for nine of them. Two (*Thermotoga maritima* and *Thermoanaerobacter tengcongensis*) have OGTs close to the threshold of 80 °C we somewhat arbitrarily used to define hyperthermophile organisms. It turns out that they both exhibit PCA2, CvP-bias, and J_2 values that could all allow their classification in continuity with previously defined hyperthermophiles.

This clear discrimination breaks down if we use an OGT of 75 °C as the hyperthermophilicity threshold. Down to this temperature, the value along the PCA2 coordinate remains a valid criterion, allowing two *Sulfolobus* species (OGT of 80 and 78 °C) to be classified in continuity with the other hyperthermophiles (PCA2 < -1.6 for OGT higher than 75 °C, and

PCA2 > -1.3 OGT lower than 75 °C; see Fig. 3c). However, the straightforward, more biochemically meaningful CvP-bias criteria become invalidated by three mesophiles with OGT of 37 °C, *Fusobacterium nucleatum* (CvP = 9.06), *Halobacterium sp* (CvP = 9.17), *Clostridium perfringens* (CvP = 9.39), exhibiting larger values than *Sulfolobus* (CvP = 8.90 and 8.80).

DISCUSSION

Using a much larger data set including many new thermophile and mesophile whole genome sequences, this follow-up study confirms the previous suggestion that the global replacement of polar residues (Asn, Gln, Ser, Thr) by charged residues (Asp, Glu, Lys, Arg) is the dominant proteome characteristic of microorganisms adapted to hyperthermophilic growth condition (OGT > 80 °C). This effect is observed for both bacteria and archaeobacteria and thus is not a simple consequence of phylogenetic relationship.

Even though the strict correspondence between the highest CvP-bias and the highest OGT breaks down below 80 °C, this property globally remains a characteristic of all thermophilic (OGT > 55 °C) microorganisms, as shown in Fig. 1c where the CvP-bias averaged over all mesothermophiles remains markedly higher than for mesophile organisms. The influence of other adaptive strategies (e.g. to high salinity or extreme pH environments), together with the phylogenetic affinities of certain mesophiles to thermophile organisms (7), probably contributes to weaken CvP-bias signal, allowing a few false positives to sneak in (such as *Clostridium perfringens* or *Streptomyces coelicolor*) (Table I.).

This study confirms that a strong CvP-bias is specifically associated with hyperthermophilic proteomes. This observation is consistent with the thermodynamic advantage resulting from the increased significance of coulomb interaction with the increasing temperature (as the dielectric constant of water decreases). The simultaneous increase of oppositely charged residues (mostly Arg, Lys, and Glu) further allows for more ion pairs to be formed at the surface of hyperthermostable proteins (2). This rationale involving the stability of proteins in a high temperature aqueous environment is also supported by our observation that the proteins predicted to be associated to the membrane (and thus designed for hydrophobic environments) exhibit a much less significant CvP-bias (data not shown). Finally, the fact that a large number of diverse genomes confirms a statistical trend previously inferred from a much smaller data set argues that increased ion-pair formation is both a significant physico-chemical factor and the preferred evolutionary pathway toward thermostable soluble proteins.

REFERENCES

1. Kumar, S., and Nussinov, R. (2001) *Cell. Mol. Life Sci.* **58**, 1216–1233
2. Cambillau, C., and Claverie, J. M. (2000) *J. Biol. Chem.* **275**, 32383–32386
3. Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K., and Suzuki, M. (2000) *Proc Natl. Acad. Sci.* **97**, 14257–14262
4. Kreil, D. P., and Ouzounis, C. A. (2001) *Nucleic Acids Res.* **29**, 1608–1615
5. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
6. Vieille, C., Epting, K. L., Kelly, R. M., and Zeikus, J. G. (2001) *Eur. J. Biochem.* **268**, 6291–6301
7. Brochier, C., and Philippe, H. (2002) *Nature* **417**, 244