Ancestral mutations as a tool for solubilizing proteins: The case of a hydrophobic phosphate-binding protein

Daniel Gonzalez, Julien Hiblot, Nune Darbinian, Jernelle C. Miller, Guillaume Gotthard, Shohreh Amini, Eric Chabriere, Mikael Elias

Ancestral mutations as a tool for solubilizing proteins: The case of a hydrophobic phosphate-binding protein☆

Daniel Gonzalez[a], Julien Hiblot[a], Nune Darbinian[b], Jernelle C. Miller[b,c], Guillaume Gotthard[a], Shohreh Amini[b,c], Eric Chabriere[a,*], Mikael Elias[d,*]

[a]URMITE UMR CNRS-IRD 6236, IFR48, Faculté de Médecine et de Pharmacie, Université de la Méditerranée, Marseille, France
[b]Department of Neuroscience, Temple University School of Medicine, Philadelphia, PA 19140, USA
[c]Department of Biology, College of Science and Technology, Temple University, Philadelphia, PA 19122, USA
[d]Weizmann Institute of Science, Biological Chemistry, Rehovot, Israel

ARTICLE INFO

ABSTRACT

Stable and soluble proteins are ideal candidates for functional and structural studies. Unfortunately, some proteins or enzymes can be difficult to isolate, being sometimes poorly expressed in heterologous systems, insoluble and/or unstable. Numerous methods have been developed to address these issues, from the screening of various expression systems to the modification of the target protein itself. Here we use a hydrophobic, aggregation-prone, phosphate-binding protein (HPBP) as a case study. We describe a simple and fast method that selectively uses ancestral mutations to generate a soluble, stable and functional variant of the target protein, here named sHPBP. This variant is highly expressed in *Escherichia coli*, is easily purified and its structure was solved at much higher resolution than its wild-type progenitor (1.3 versus 1.9 Å, respectively).

© 2013 The Authors. Published by Elsevier B.V. on behalf of Federation of European Biochemical Societies. All rights reserved.

Introduction

Stable, soluble and functional proteins comprise ideal models for functional and structural studies. However, when overexpressed in heterologous systems such as in *Escherichia coli*, natural proteins from various sources can sometimes be insoluble, unstable or poorly expressed [1]. These difficulties considerably hamper studies of certain proteins, and have yielded a considerable bias in protein functional and structural analysis toward soluble and expressible proteins [2]. Several strategies have been developed to skirt these limitations. Classical methods involving expression in heterologous systems usually screen the host nature, the culture conditions, and media composition [1,3–5]. Codon optimization, protein fusion or the co-expression with chaperones [6] may also represent useful strategies to successfully express proteins [1,5,6]. Nevertheless, these trials may remain inefficient in some cases; particularly for numerous mammalian proteins [1]. Thus, methodologies tuning the protein target itself emerged,

with the aim of producing soluble and expressible models for further studies [7,8]. In that respect, site-directed mutagenesis could be used to substitute surface residues and therefore generate more soluble proteins. However, this technique is limited by our ability to precisely identify problematic surface residues [9]. Conversely, directed evolution allows to extensively mutate the target protein and to select for more soluble variants [10]. The explored sequence space is however huge, and the method thus requires a high throughput screening method [11]. An alternative method, called DNA shuffling, uses several genes sharing high sequence identity with the target protein and shuffles them all. The screening of the resulting gene library for solubility may yield soluble and expressible variants, which can subsequently be subjected to functional and structural studies [7].

In regard to protein stabilization or solubilization, phylogenetic-based protein engineering may represent a powerful method. Indeed, consensus libraries, which are composed of mutations that bring the sequence of the target gene closer to the family consensus sequence, can efficiently stabilize proteins [12]. Ancestral mutations also have proven ability to yield soluble and stabilized protein variants [13,14]. Moreover, the use of ancestral mutations libraries can yield interesting protein variants with altered enzymatic activity and/or stability [11,15–18]. Ancestral mutations may therefore be turned into an easy and fast method to solubilize/stabilize contemporary proteins.

In this study, we focus on ancestral mutations, and their ability to solubilize a protein target. We therefore used the human phosphate binding protein (HPBP) as a case study. HPBP belongs to DING proteins

family, a clade of the phosphate binding protein (PBP) superfamily [19]. HPBP is a hydrophobic, possibly an apolipoprotein, crystallized from supposedly pure human paraoxonase (PON1) preparations [20–22]. HPBP possesses a venus-flytrap topology identical to the high affinity phosphate-binding protein (PBP or PstS) carriers of the ABC transporter systems, and a similar phosphate-binding ability [22]. Interestingly, and as for other related DING proteins [23–25], HPBP has been shown to inhibit HIV-1 replication by targeting the transcriptional step [26].

Nevertheless, functional and structural studies on HPBP are considerably hampered by its high hydrophobicity [22], its propensity to aggregate, and the failure to express it heterologously in soluble form (Chabriere, unpublished results). The existing purification procedure, starting from human plasma samples, is complex and laborious, and yields to little amounts of pure HPBP [20]. We hereby describe a simple and fast method to generate an ancestral-mutations based, fully functional, soluble variant of HPBP.

Materials and methods

Phylogenetic analysis and ancestral resurrection

The sequences of phosphate-binding proteins (including DING and PstS proteins) were collected from the National Center of Biotechnology Information (NCBI) using protein alignment BLAST (blastp) [27,28] with default parameters *versus* the non-redundant protein sequence database (nr). Only complete protein sequences were selected, and redundancy was subsequently removed (maximum 95% of sequence identity) with Cd-hit [29]. The sequence alignment was performed with clustalW 2.0 software [30] and manually improved (Fig. 1-1). The substitution matrix corresponding to the sequence alignment was determined using the Prottest software [31]. The alignment was subsequently submitted to PhyML software [32] with the JTT substitution matrix with 100 iterations. The prediction of the putative ancestral sequences at each nodes was performed using FastML [33]. We have chosen the putative ancestor of HPBP and a related, bacterial, soluble homologue, PfluDING from *Pseudomonas fluorescens* (Node 10) (Fig. 1-2) [34,35]. The ancestral mutation library contains the substitutions of the putative ancestral sequence, as compared to HPBP sequence. By applying four simple filters:

(1) include substitutions of surface apolar residue into polar residue,
(2) include substitution from Gly to X, with the exception of Gly residue involved in the start/end of secondary structure,
(3) include core mutations,
(4) include mutations that change a surface hydrophobic residue into a less hydrophobic one, we have reduced the numbers of substitutions from 93 to 22 (Fig. 1-3).

Gene synthesis and cloning of HPBP and sHPBP

The genes encoding for HPBP and sHPBP were optimized for *E. coli* expression and synthesized by service providers (Genecust, Luxembourg, and GeneArt, Life Technologies, France, respectively) (Fig. 1-4). The genes were subsequently subcloned into pET22b (+) (Novagen) using *Nco*I and *Xho*I as cloning sites.

Production and purification of HPBP and sHPBP

Productions of HPBP and sHPBP were performed using *E. coli* BL21(DE3)-pGro7/GroEL cells (TaKaRa) in 6 l of ZYP medium [1] (100 μg ml^{-1} ampicillin, 34 μg ml^{-1} chloramphenicol). The cultures were grown at 37 °C to reach OD$_{600nm}$ = 0.6 and then induced by starting the consumption of lactose in ZYP medium coupled to temperature transition to 17 °C during 16 h. Cells were harvested by

centrifugation (4500g, 4 °C, 15 min) and pellets were suspended in 400 ml of lysis buffer (20 mM TRIS, pH 8, 100 mM NaCl, Lysozyme 0.25 mg ml^{-1}, DNAse I 10 μg ml^{-1}, PMSF 0.1 mM, MgSO$_4$ 20 mM and 8 tablets of anti-protease EDTA-free (Roche)) and stored at −80 °C for 2 h. Cells were then thawed at 37 °C for 15 min and disrupted by 3 steps of 30 s of sonication (QSonica sonicator; amplitude 40). Debris was removed by centrifugation (12,500g, 4°C, 30 min). Supernatant was loaded on a Nickel affinity column (HisTrap 5 ml, FFCrude from GE Healthcare) at a flow rate of 5 ml min^{-1}. Proteins gripped to the column were eluted by imidazol, using an elution buffer (20 mM TRIS, pH 8, 100 mM NaCl and 250 mM imidazole). Then, a size exclusion chromatography step (Superdex 75 16/60, GE Healthcare) was performed using buffer 20 mM TRIS, pH 8 and 100 mM NaCl. Protein production and purity were checked by 15% SDS–PAGE analysis (Fig. 1-5) and mass spectrometry analysis (Plateforme Timone, Marseille, France).

Crystallization of sHPBP

sHPBP was concentrated to 4 mg ml^{-1} using a centrifugation device (Vivaspin 500, MWCO 3 kDa, Sartorius stedim, Germany). Crystallization trials were performed at 298 K using the same condition as the homologue PfluDING (*i.e.* LiSO$_4$ 1 M, 20–30% PEG 8000 and Sodium Acetate 2 M at pH 4.5–5.5, [34,36,37]). Since only thin crystal plates were obtained, commercial screens conditions were tested (*i.e.* Stura and MDL, Molecular Dimension, England) using a Mosquito instrument (TTP Labtech, England) with the sitting-drop vapor diffusion method setup in a 96-well plate. Drops were monitored using a Discovery V8 binocular microscope and an AxioCam ERc5S camera (Zeiss, Germany). Crystals were obtained in the MDL screen, in a condition containing 0.2 M Sodium Acetate, 0.1 M Sodium Cacodylate pH 6.5 and 25% PEG 8000, and using a 2:1 (protein:reservoir) ratio (200 nl:100 nl). In order to obtain bigger crystals, this condition was optimized using the hanging drop method. The final condition (0.2 M Sodium Acetate, 0.1 M Sodium Cacodylate pH 6.5 and 25% PEG 8000), and using a 2:1 (protein:precipitant) ratio (500 nl:250 nl), led to the apparition of three-dimensional crystals (around 75–100 μm). Reproducible crystals appeared after 3 days at 298 K.

Data collection and structure resolution of sHPBP

The crystal was transferred few seconds in a drop (1 μl) containing a cryo-protectant solution made out of the mother liquor plus 10% (v/v) of glycerol. After mounting on a CryoLoop (Hampton research), crystal was flash-frozen in liquid nitrogen. X-ray diffraction intensities were collected on the ID23-1 beamline at the ESRF (Grenoble, France) using a wavelength of 0.97655 Å and a Pilatus 6M detector with 0.15 s exposures. Diffraction data were collected from 1027 images; each frames consisted of 0.15° step oscillations, over a range of 154.05° (Table 1). The molecular replacement using the HPBP structure as model (PDB: 2V3Q) was performed with *Phaser* [38]. The solution was then used for refinement performed using *REFMAC5* and *phenix* [39,40], the model was improved using *Coot* [41]. The model and structure factor were deposited under the Protein Data Bank code PDB: 4M1V (Table 1).

HIV inhibition by HPBP and sHPBP

Inhibition assays were performed in HeLa cells, as previously described [23]. Briefly, HeLa cells were maintained in DMEM + 10 % FBS with antibiotics (100 U/ml of Penicillin and 100 mg/ml of Streptomycin). Cells were pre-incubated with proteins (0.25 μg/ml) for 48 h, then transfected using Lipofectamine 2000 transfection reagent with 0.5 μg of HIV-LTR-luciferase reporter plasmid. Each transfection was done in triplicate.
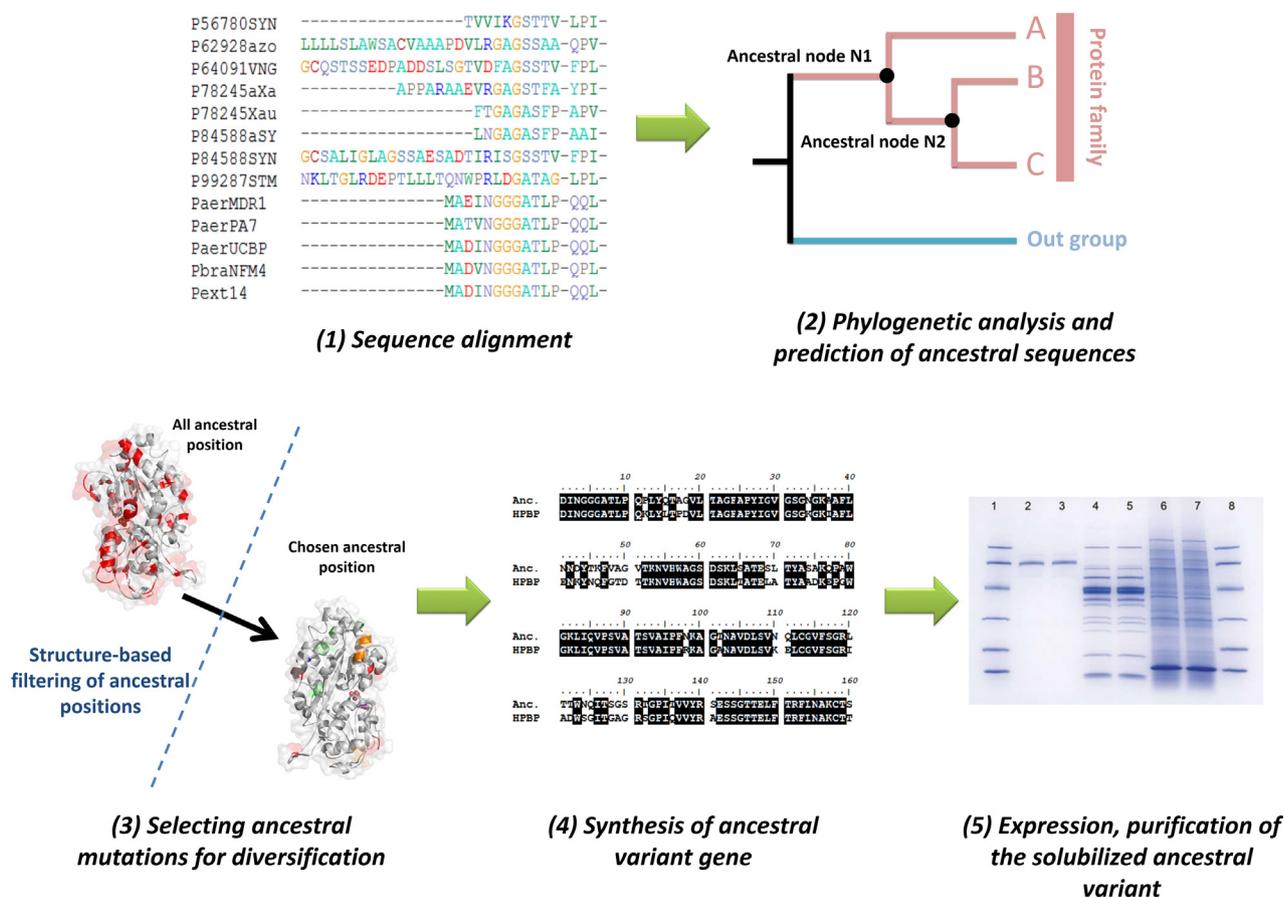
(1) Sequence alignment

(2) Phylogenetic analysis and prediction of ancestral sequences

(3) Selecting ancestral mutations for diversification

(4) Synthesis of ancestral variant gene

(5) Expression, purification of the solubilized ancestral variant

Fig. 1. Selective use of ancestral mutations for protein solubilization. (1) Orthologous and paralogous protein sequences of the target protein has been identified and aligned. (2) A phylogenetic tree was constructed using this alignment, and the most probable ancestral sequence at each node were inferred. The substitutions between the most probable ancestral sequence at a given node and the contemporary target protein sequence (*i.e wt* HPBP) delineate the ancestral library. (3) Using a structural model and simple criteria (see methods), ancestral positions were filtered. (4) The gene coding for the ancestral variant was synthetized and (5) the protein was expressed, purified and crystallized.

Table 1
Data collection and refinement statistics of sHPBP structure.

Data collection	
Dataset	Native
PDB ID	4M1V
Beamline	ID23-1
Wavelength (Å)	0.97655
Detector	PILATUS 6M
Oscillation (°)	0.15
Number of frames	1027
Resolution (Å) (last bin)	1.3 (1.4–1.3)
Space group	C2
Unit-cell parameters (Å)	$a = 125.04, b = 71.99, c = 38.98, \beta = 103.12$
No. of observed reflections (last bin)	268471 (52411)
No. of unique reflections (last bin)	81625 (16226)
Completeness (%)(last bin)	98.8 (99.3)
R_{meas} (%) (last bin)	4.1 (27.5)
CC (1/2) (last bin)	99.9 (95.3)
$I/\sigma(I)$ (last bin)	21.36 (4.65)
Redundancy (last bin)	3.28 (3.23)
Refinement statistics	
R_{free}/R_{work}	13.96/10.47
No. of total model atoms	3541
Ramachandran favored (%)	98.4
Ramachandran outliers (%)	0
Generously allowed rotamers (%)	1.6
Rmsd from ideal	
Bond lengths (Å)	0.027
Bond angles (°)	2.466

Results

Ancestral mutations as a tool for protein solubilization

A phylogenetic tree of the PBPs, including DING proteins was generated (Figs. 1 and 2A), and the most probable ancestral sequence at Node 10 was reconstructed (Fig. 2A, see methods). The sequence of this putative ancestor exhibits 93 substitutions with the contemporary HPBP sequence (Fig. 2B). Among them, 22 substitutions were rationally selected (Fig. 2B and C). The selection was made using very simple criteria, such as (i) include substitutions of surface apolar/hydrophobic residues into polar/less hydrophobic ones, (ii) include core mutations, and (iii) include substitutions of glycine to X, when the glycine residue is not involved in the start/end of a secondary structure (see methods). The selected substitutions are in fact mainly located at the protein surface (Fig. 2D), and are predicted to solubilize the target protein, as shown by the hydrophobic profile comparison of sHPBP and HPBP (Fig. 3).

The gene coding for sHPBP, the ancestral variant that includes 22 substitutions, was synthesized, and both sHPBP and *wt* HPBP were heterologously expressed in *E. coli* strain BL21(DE3)-pGro7/EL, using the same protocol. Protein expression profiles show that *wt* HPBP is expressed in *E. coli*, albeit solely in the insoluble fraction (Fig. 4, left part). Conversely, sHPBP is largely present in the soluble fraction (Fig. 4, middle part). Noteworthy, both proteins share more than 94% of sequence identity but possess opposite solubility profiles in *E. coli*. The expression of sHPBP is significant, as illustrated by the purification yield: about 12 mg of pure protein per liter of culture (Fig. 4,
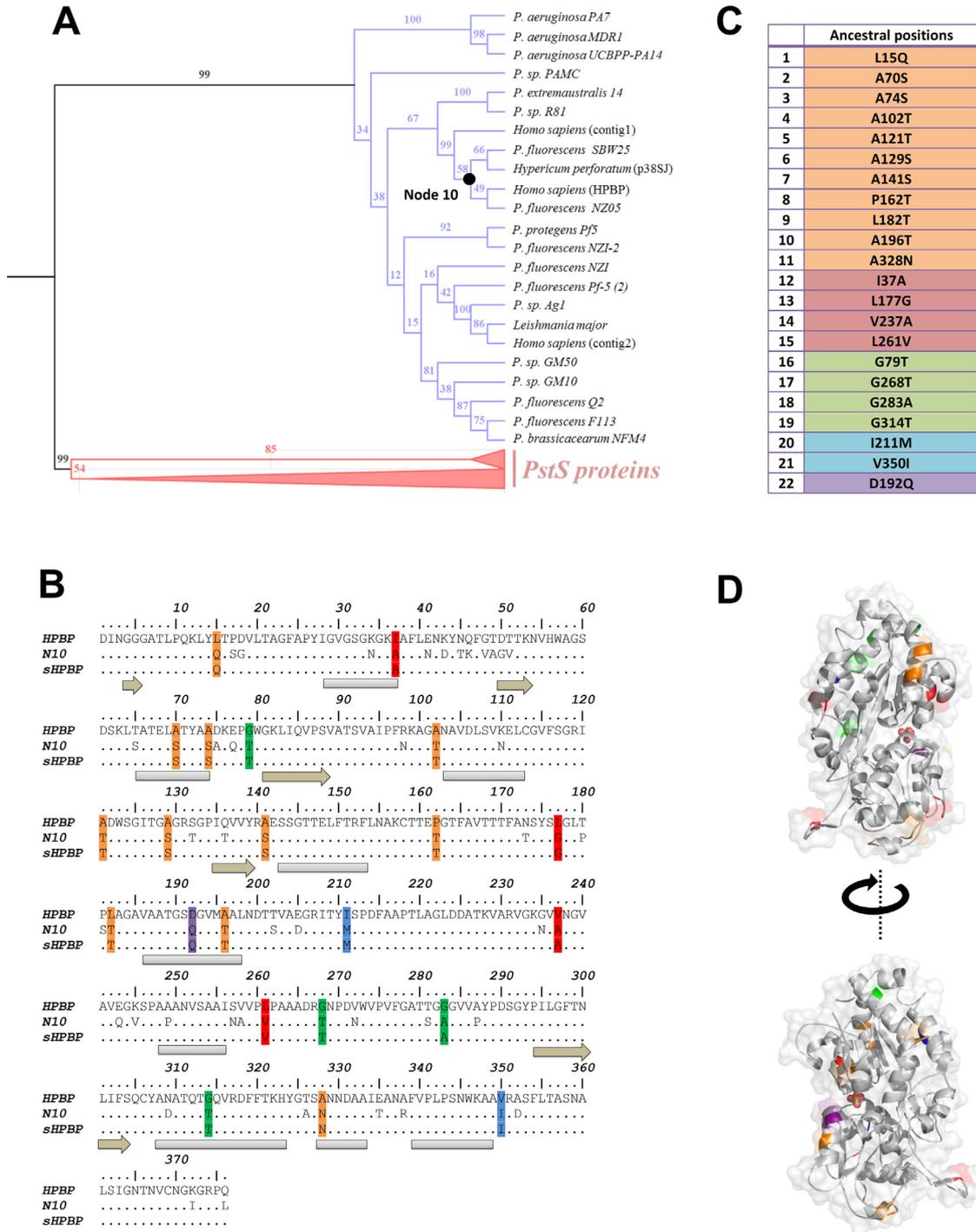
A

```
                                                        P. aeruginosa PA7
                                              100
                                                 98   P. aeruginosa MDR1
                                                       P. aeruginosa UCBPP-PA14
                           99                          P. sp. PAMC
                                                100    P. extremaustralis 14
                                       34    67         P. sp. R81
                                                   99   Homo sapiens (contig1)
                                                    66   P. fluorescens SBW25
                                                    58    Hypericum perforatum (p38SJ)
                                           38   Node 10  49  Homo sapiens (HPBP)
                                                          P. fluorescens NZ05
                                                    92    P. protegens Pf5
                                                          P. fluorescens NZI-2
                                               12   16   P. fluorescens NZI
                                                    42    P. fluorescens Pf-5 (2)
                                               15   100   P. sp. Ag1
                                                     86   Leishmania major
                                                          Homo sapiens (contig2)
                                                     81   P. sp. GM50
                                                     38   P. sp. GM10
                                                       87  P. fluorescens Q2
                                                        75  P. fluorescens F113
                                                           P. brassicacearum NFM4
                           99      85
                              54                        PstS proteins
```

C

	Ancestral positions
1	L15Q
2	A70S
3	A74S
4	A102T
5	A121T
6	A129S
7	A141S
8	P162T
9	L182T
10	A196T
11	A328N
12	I37A
13	L177G
14	V237A
15	L261V
16	G79T
17	G268T
18	G283A
19	G314T
20	I211M
21	V350I
22	D192Q

B

```
              10        20        30        40        50        60
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     DINGGGATLPQKLYLTPDVLTAGFAPYIGVGSGKGKAFLENKYNQFGTDTTKNVHWAGS
N10      ..............Q.SG........N.....N.D.TK.VAGV....
sHPBP    ..............Q..........................

              70        80        90       100       110       120
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     DSKLTATELATYAADKEPGWGKLIQVPSVATSVAIPFRKAGANAVDLSVKELCGVFSGRI
N10      ...S.....S....SA.Q.I...........N..T......N.....
sHPBP    .........S....S..I.................T......

             130       140       150       160       170       180
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     ADWSGITGAGRSGPIQVVYRAESSGTTELFTRFLNAKCTTEPGTFAVTTTFANSYSIGLT
N10      T.......S..T...T...S...........T.......T..I..P
sHPBP    T.........S...S................T..........I..

             190       200       210       220       230       240
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     PLAGAVAATGSDGVMALNDTTVAEGRITYISPDFAAPTLAGLDDATKVARVGKGVVNGV
N10      ST........Q..T.....S.D........M.................NA..A.
sHPBP    .T.........Q...................M..................A.

             250       260       270       280       290       300
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     AVEGKSPAAANVSAAISVVPIPAAADNSNPDVWVPVFGATTCGVVAYPDSGYPILGFTN
N10      ..Q.V...P.....NA....N........S..A..P..........
sHPBP    ........................................

             310       320       330       340       350       360
         ....|....|....|....|....|....|....|....|....|....|....|....|
HPBP     LIFSQCYANATQTGQVRDFFTKHYGTSANNDAAIEANAFVPLPSNWKAAVRASFLTASNA
N10      ........D....G.........A.N.......T..R.......I.D...
sHPBP    ...................................I..T...

             370
         ....|....|....|.
HPBP     LSIGNTNVCNGKGRPQ
N10      ..........I...L
sHPBP    ...............
```

D

Fig. 2. Phylogeny of PBPs and selected ancestral mutations. (A) Phylogenetic tree of PBPs, including PstS proteins (in pink) and DING proteins (in blue). Bootstrap values are shown for each node. (B) Sequence alignment of the target protein sequence (*wt* HPBP), the most probable ancestor sequence (Node 10) and the ancestral variant sequence containing selected substitutions (sHPBP). (C) Ancestral substitutions that were retained after structure-based filtering: substitutions changing apolar residues into polar residues (orange), hydrophobic residues into less hydrophobic ones (magenta), core substitutions (cyan) and D192Q substitution (purple) are listed. Secondary structures are represented in gray and light brown, for α-helix and β-sheet, respectively. (D) Retained ancestral substitutions locations on sHPBP structure are colored as in (C). The protein surface is represented as transparent gray and the bound phosphate anion is shown as red spheres. (For interpretation of the references to colour in this figure legend, the reader is reffered to the web version of this article.)
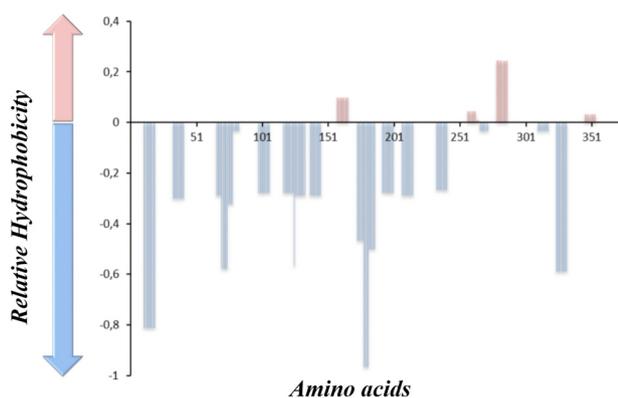
Fig. 3. Hydropathy profile discrepancy between sHPBP and HPBP. The hydropathy profiles of both proteins were determined using the software "ProtScale" in Expasy website (algorithm Kyte and Doolitle, Window size 5). The graph represents, in y-axis, the difference of the hydrophobic potential between sHPBP and HPBP along the protein sequence (in x-axis). The light blue bars are for the sequence regions where sHPBP is predicted to be the more hydrophilic than HPBP, whereas regions where sHPBP is predicted to be more hydrophobic are shown as pink bars. (For interpretation of the references to colour in this figure legend, the reader is reffered to the web version of this article.)
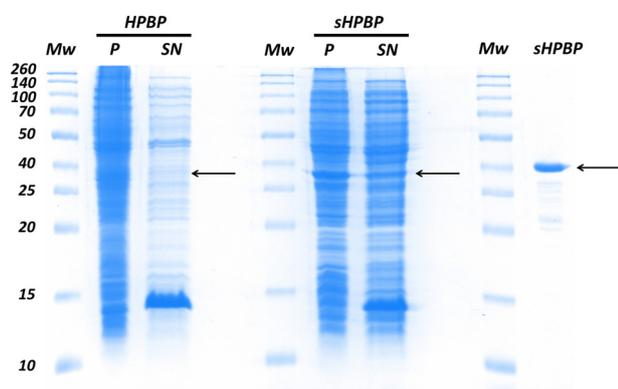
Fig. 4. Solubility of *wt* HPBP and sHPBP when overexpressed in *E. coli*. The *left part, middle part* and *left part* of this Coomassie-stained SDS–PAGE gel correspond to the expression of HPBP, the expression of sHPBP, and the purified sHPBP, respectively. The lanes "P" and "SN" are for pellet and supernatant, respectively. Molecular weight markers (lane "Mw") are indicated in kiloDalton (kDa; Spectra Multicolor broad range protein ladder). Arrows indicate the expected position of HPBP and sHPBP.

right part). Additionally, whereas HPBP is aggregation-prone, sHPBP is soluble, including in the absence of detergent, and present a melting temperature (T_m) of $47.2 \pm 0.13\,^\circ\text{C}$.

sHPBP and wt HPBP share nearly identical structures

sHPBP yielded high quality crystals and its structure was solved at 1.3 Å resolution (Table 1), a resolution that is higher than the previous structure of HPBP (1.9 Å) [42]. As expected by the high sequence identity between both proteins (94% identity), the structure of sHPBP is extremely similar to that of HPBP. The structures of both proteins exhibit identical topology, where the two globular domains are linked together by a flexible hinge. Each domain is composed of a central core β-sheets which is flanked by α-helix and present at their interface the phosphate binding cleft. Both structures superimpose extremely well, with a RMSD on all carbon α positions of 0.2 Å (Fig. 5A), including in the phosphate binding cleft region (Fig. 5B). Nonetheless, small differences relate to slightly different conformations of the 4 protuberant surface loops, some being possibly affected by ancestral mutations. The crystal packings of HPBP and sHPBP being different, we cannot exclude that packings affected the conformations of these

surface loops. Interestingly, the calculation of the surface electrostatic potential of both structures reveals significantly different patterns, consistent with a superior solubility of sHPBP, since the apolar (or hydrophobic) surface patches present in HPBP structure are smaller in the sHPBP structure (Fig. 5C).

The soluble variant sHPBP is functional

Besides the phosphate binding ability, HPBP has been shown to exhibit anti-HIV-1 properties [26]. The anti-HIV properties of HPBP and sHPBP have thus been compared using a LTR-reporter system in HeLa cells. Each assay was performed by pre-incubation of proteins at the estimated IC_{50} value of HPBP (*i.e.* 5 nM), previously determined using derived-immune system cells (Peripheral blood lymphocytes and primary macrophages) [26]. At this concentration and on HeLa cells, HPBP has a strong inhibitory effect on the HIV transcription (78%). We here show that sHPBP inhibits the HIV transcription with a very similar efficiency (68%) (Fig. 5D). This strongly suggests that sHPBP shares similar inhibition properties (*e.g.* IC_{50}, CC_{50}) and mechanisms than *wt* HPBP.

Discussion

Ancestral variants such as sHPBP comprise excellent model for functional studies of poorly soluble proteins

HPBP is a protein with an increasing interest because of its HIV-1 inhibition properties. However, its study has long been hampered because of its intrinsic hydrophobic character, and the resulting difficulty to obtain and store purified protein. We used ancestral mutations to generate a soluble variant of HPBP. This variant, sHPBP, can be expressed and purified from a convenient host, *E. coli*, with a significant purification yield (~12 mg/l of culture), whereas this classical approach failed with *wt* HPBP. The enhanced expression level observed for sHPBP is concomitant to its higher solubility, compared to the *wt* protein. This dramatic change in behavior may be mainly ascribed to the decrease of the protein's surface hydrophobicity. sHPBP is therefore a more convenient protein to work with, and a critical consequence of this fact resides in the obtaining of sHPBP structure at 1.3 Å resolution; whereas HPBP structure was solved at 1.9 Å. Indeed, a better control on the purification process and the availability of high amounts of purified protein are key factors for optimization of crystallization conditions.

The structure of sHPBP reveals that it is extremely similar to that of *wt* HPBP. Most importantly, the phosphate binding cleft of both proteins is nearly identical. Consequently to this high similarity, sHPBP shares a similar HIV-1 inhibition capacity with *wt* HPBP. sHPBP therefore represents an excellent model for future structural and functional studies, with the aim of deciphering the biological function(s) of HPBP.

A fast, easy and efficient method based on ancestral mutations can produce soluble variants

We describe here the construction of a soluble variant of the hydrophobic, aggregation-prone, HPBP using ancestral mutations. The employed methodology is fast, simple and yielded a variant that is both soluble and active. Ancestral mutations and libraries indeed comprise an efficient tool for focusing substitutions to positions that can readily promote changes in protein stability, solubility or even substrate specificity [11,13,43]. A primordial property of ancestral mutations resides in the possibility of incorporating a very large number of these mutations, while maintaining the produced variants viable [11,43]. Therefore, by properly choosing ancestral mutations (*e.g.* active site substitutions), one can efficiently alter the substrate specificity of enzymes [11].
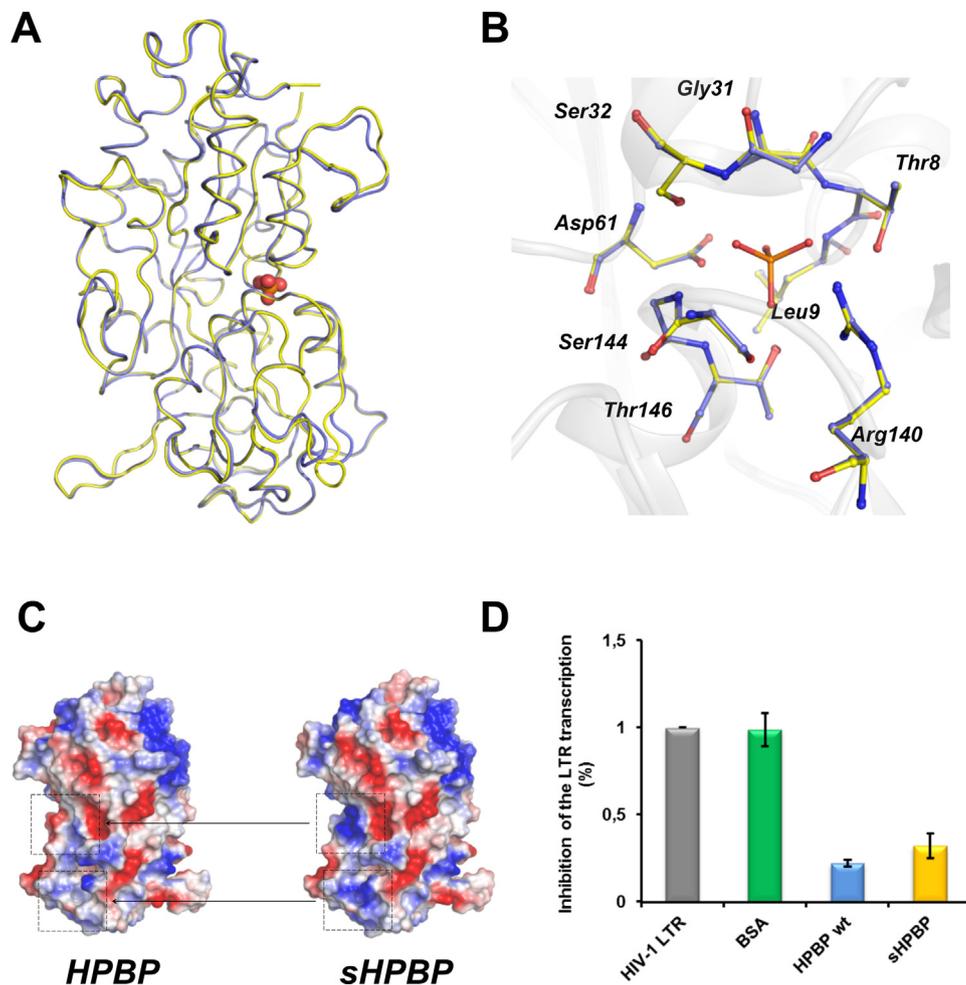
Fig. 5. Structural and functional comparison of sHPBP and *wt* HPBP. (A) Structural superposition of sHPBP (yellow) and *wt* HPBP (blue). Phosphate is represented as red spheres. (B) Superposition of the phosphate binding cleft of both proteins (sHPBP (yellow) and HPBP (blue)). (C) Electrostatic potential comparison of both proteins. Positive and negative charges are respectively represented in red and blue. White regions are neutrally charged or hydrophobic. (D) Inhibition of the HIV-1 transcription, measured using a LTR-luciferase reporter system in HeLa cells. The constitutive luminescence of the system is represented in HIV-1 LTR (gray). The inhibition properties of the different samples (0.25 µg/ml) are shown: Bovine Serum Albumin (green), *wt* HPBP (blue) and sHPBP (yellow). (For interpretation of the references to colour in this figure legend, the reader is reffered to the web version of this article.)

Here, we chose ancestral node 10 that relates the hydrophobic HPBP to a bacterial homologue dubbed PfluDING, reported to be a soluble and stable protein [34]. Then, among the substitutions between the putative ancestors and the HPBP sequence, we have mainly retained surface residues that may have a stabilizing and/or a solubilizing effect by applying 4 simple criteria (see Results). The resulting sequence was subsequently synthetized and yielded a soluble variant, sHPBP. The requirements for this method are limited to a good sequence sampling of the protein family, as well as a structure model for proper selection of the inferred ancestral mutations. This method can therefore be applied in numerous cases.

We believe that the selective use of ancestral mutations described here complements the existing approaches for producing more stable proteins, such as directed evolution, family shuffling, consensus libraries and others. However, its unique feature and its simplicity make it attractive to use with challenging protein targets, with the aim of producing, stable, soluble variants with increased propensity to crystallize.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fob.2013.12.006.

References

[1] Graslund, S, Nordlund, P, Weigelt, J, Hallberg, BM, Bray, J, Gileadi, O et al. (2008) Protein production and purification. Nat. Methods 5, 135–146.
[2] Peng, K, Obradovic, Z, Vucetic, S. (2003). Exploring bias in the Protein Data Bank using contrast classifiers. in: Paper Presented at the *Pacific Symposium on Biocomputing 2004*: Hawaii, USA, pp. 6–10, January 2004.
[3] Sorensen, HP and Mortensen, KK (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb. Cell Fact. 4, 1.
[4] Sorensen, HP and Mortensen, KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J. Biotechnol. 115, 113–128.
[5] Vincentelli, R, Bignon, C, Gruez, A, Canaan, S, Sulzenbacher, G, Tegoni, M et al. (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. Acc. Chem. Res. 36, 165–172.
[6] Tokuriki, N and Tawfik, DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459, 668–673.

[7] Harel, M, Aharoni, A, Gaidukov, L, Brumshtein, B, Khersonsky, O, Meged, R et al. (2004) Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. Nat. Struct. Mol. Biol. 11, 412–419.

[8] Aharoni, A, Gaidukov, L, Yagur, S, Toker, L, Silman, I and Tawfik, DS (2004) Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. Proc. Natl. Acad. Sci. U.S.A. 101, 482–487.

[9] Miyazaki, K and Arnold, FH (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J. Mol. Evol. 49, 716–720.

[10] Waldo, GS (2003) Genetic screens and directed evolution for protein solubility. Curr. Opin. Chem. Biol. 7, 33–38.

[11] Alcolombri, U, Elias, M and Tawfik, DS (2011) Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. J. Mol. Biol. 411, 837–853.

[12] Jochens, H, Aerts, D and Bornscheuer, UT (2010) Thermostabilization of an esterase by alignment-guided focussed directed evolution. Protein Eng. Des. Sel. 23, 903–909.

[13] Gaucher, EA, Govindarajan, S and Ganesh, OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451, 704–707.

[14] Risso, VA, Gavira, JA, Mejia-Carmona, DF, Gaucher, EA and Sanchez-Ruiz, JM (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. J. Am. Chem. Soc. 135, 2899–2902.

[15] Afriat-Jurnou, L, Jackson, CJ and Tawfik, DS (2012) Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling Biochemistry.

[16] Khersonsky, O, Rosenblat, M, Toker, L, Yacobson, S, Hugenmatter, A, Silman, I et al. (2009) Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization. Biochemistry. 48, 6644–6654.

[17] Bershtein, S, Goldin, K and Tawfik, DS (2008) Intense neutral drifts yield robust and evolvable consensus proteins. J. Mol. Biol. 379, 1029–1044.

[18] Jochens, H, Aerts, D and Bornscheuer, UT (2010) Thermostabilization of an esterase by alignment-guided focussed directed evolution. Protein Eng. Des. Sel. 23, 903–909.

[19] Berna, A, Scott, K, Chabriere, E and Bernier, F (2009) The DING family of proteins: ubiquitous in eukaryotes, but where are the genes? Bioessays 31, 570–580.

[20] Renault, F, Chabriere, E, Andrieu, JP, Dublet, B, Masson, P and Rochu, D (2006) Tandem purification of two HDL-associated partner proteins in human plasma, paraoxonase (PON1) and phosphate binding protein (HPBP) using hydroxyapatite chromatography. J. Chromatogr. 836, 15–21.

[21] Rochu, D, Chabriere, E, Elias, M, Renault, F, Clery-Barraud, C and Masson, P (2008) Stabilisation of active form of natural human PON1 requires HPBP. The Paraoxonases: Their Role in Disease Development and Xenobiotic Metabolism. Netherlands: Springer, pp. 171–183.

[22] Morales, R, Berna, A, Carpentier, P, Contreras-Martel, C, Renault, F, Nicodeme, M et al. (2006) Serendipitous discovery and X-ray structure of a human phosphate binding apolipoprotein. Structure 14, 601–609.

[23] Darbinian, N, Gomberg, R, Mullen, L, Garcia, S, White, MK, Khalili, K, Amini, S. Suppression of HIV-1 transcriptional elongation by a DING phosphatase J. Cell. Biochem. 112, 225–232.

[24] Sachdeva, Rakhee, Darbinian, N, Khalili, Kamel, Amini, Shohreh, Gonzalez, Daniel, Djeghader, Ahmed et al. (2013) DING proteins from phylogenetically different species share high degree of sequence and structure homology and block transcription of HIV-1 LTR promoter PloS One.

[25] Lesner, A, Shilpi, R, Ivanova, A, Gawinowicz, MA, Lesniak, J, Nikolov, D et al. (2009) Identification of X-DING-CD4, a new member of human DING protein family that is secreted by HIV-1 resistant CD4(+) T cells and has anti-viral activity. Biochem. Biophys. Res. Commun. 389, 284–289.

[26] Cherrier, T, Elias, M, Jeudy, A, Gotthard, G, Le Douce, V, Hallay, H et al. (2011) uman-phosphate-binding-protein inhibits HIV-1 gene transcription and replication. Virol. J. 8, 352.

[27] Altschul, SF, Gish, W, Miller, W, Myers, EW and Lipman, DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

[28] Mount, DW (2007) Using the Basic Local Alignment Search Tool (BLAST) CSH Prot., 2007, pdb top17 .

[29] Li, W and Godzik, A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

[30] Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

[31] Darriba, D, Taboada, GL, Doallo, R and Posada, D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

[32] Guindon, S and Gascuel, O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

[33] Ashkenazy, H, Penn, O, Doron-Faigenboim, A, Cohen, O, Cannarozzi, G, Zomer, O et al. (2012) FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic acids research. 40, W580–W584.

[34] Moniot, S, Elias, M, Kim, D, Scott, K and Chabriere, E (2007) Crystallization, diffraction data collection and preliminary crystallographic analysis of DING protein from *Pseudomonas fluorescens*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63, 590–592.

[35] Ahn, S, Moniot, S, Elias, M, Chabriere, E, Kim, D and Scott, K (2007) Structure–function relationships in a bacterial DING protein. FEBS Lett. 581, 3455–3460.

[36] Liebschner, D, Elias, M, Moniot, S, Fournier, B, Scott, K, Jelsch, C et al. (2009) Elucidation of the phosphate binding mode of DING proteins revealed by sub-angstrom X-ray crystallography. J. Am. Chem. Soc. 131, 7879–7886.

[37] Elias, M, Wellner, A, Goldin-Azulay, K, Chabriere, E, Vorholt, JA, Erb, TJ et al. (2012) The molecular basis of phosphate discrimination in arsenate-rich environments. Nature 491, 134–137.

[38] McCoy, AJ, Grosse-Kunstleve, RW, Adams, PD, Winn, MD, Storoni, LC and Read, RJ (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.

[39] Adams, PD, Afonine, PV, Bunkoczi, G, Chen, VB, Davis, IW, Echols, N et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221.

[40] Vagin, AA, Steiner, RA, Lebedev, AA, Potterton, L, McNicholas, S, Long, F et al. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2184–2195.

[41] Emsley, P, Lohkamp, B, Scott, WG and Cowtan, K (2010) Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 486–501.

[42] Diemer, H, Elias, M, Renault, F, Rochu, D, Contreras-Martel, C, Schaeffer, C et al. (2008) Tandem use of X-ray crystallography and mass spectrometry to obtain ab initio the complete and exact amino acids sequence of HPBP, a human 38-kDa apolipoprotein. Proteins 71, 1708–1720.

[43] Gaucher, EA, Thomson, JM, Burgan, MF and Benner, SA (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425, 285–288.