Molecular Evolution of the Tissue-nonspecific Alkaline Phosphatase Allows Prediction and Validation of Missense Mutations Responsible for Hypophosphatasia*§

Received for publication, April 25, 2014, and in revised form, July 11, 2014 Published, JBC Papers in Press, July 14, 2014, DOI 10.1074/jbc.M114.576843

Jérémie Silvent1,2, Barbara Gasse1,3, Etienne Mornet5, and Jean-Yves Sire4,§

From the 1Université Pierre & Marie Curie, IBPS, Evolution Paris Seine, 7 quai St-Bernard, Case 05, 75005 Paris and the 2Université de Pathologie Cellulaire et Génétique, EA2493, Université de Versailles-Saint Quentin en Yvelines, Versailles & Unité de Génétique Constitutionnelle, Centre Hospitalier de Versailles, 78150 Le Chesnay, France

Background: We used evolutionary analysis of alkaline phosphatase (tissue nonspecific alkaline phosphatase, TNSALP) to predict missense mutations leading to hypophosphatasia.

Results: We found 469 sensitive positions and validated 99% of the 204 known mutations.

Conclusion: It is a more powerful method than in silico models to validate missense mutations in TNSALP.

Significance: Such an approach should be widely used to support genetic diagnostics.

ALPL encodes the tissue nonspecific alkaline phosphatase (TNSALP), which removes phosphate groups from various substrates. Its function is essential for bone and tooth mineralization. In humans, ALPL mutations lead to hypophosphatasia, a genetic disorder characterized by defective bone and/or tooth mineralization. To date, 275 ALPL mutations have been reported to cause hypophosphatasia, of which 204 were simple missense mutations. Molecular evolutionary analysis has proved to be an efficient method to highlight residues important for the protein function and to predict or validate sensitive positions for genetic disease. Here we analyzed 58 mammalian TNSALP to identify amino acids unchanged, or only substituted by residues sharing similar properties, through 220 millions years of mammalian evolution. We found 469 sensitive positions of the 524 residues of human TNSALP, which indicates a highly constrained protein. Any substitution occurring at one of these positions is predicted to lead to hypophosphatasia. We tested the 204 missense mutations resulting in hypophosphatasia against our predictive chart, and validated 99% of them. Most sensitive positions were located in functionally important regions of TNSALP (active site, homodimeric interface, crown domain, calcium site, . . .). However, some important positions are located in regions, the structure and/or biological function of which are still unknown. Our chart of sensitive positions in human TNSALP enables to validate or invalidate at low cost any ALPL mutation, which would be suspected to be responsible for hypophosphatasia, by contrast with time consuming and expensive functional tests, and (ii) displays higher predictive power than in silico models of prediction.

The human alkaline phosphatase (ALP)5 family comprises four members: the tissue nonspecific (liver/bone/kidney) ALP, called TNSALP, which is encoded by ALPL, and three tissue-specific ALPs: intestinal (IAP, encoded by ALPI), placental (PLAP-1 encoded by ALPP), and placental-like 2 or germ cell (GCAP encoded by ALPPL2) (1–4). In humans, ALPL is located on chromosome 1 (1p36.1-p34), whereas the genes of the three other ALPs are clustered on chromosome 2. ALPL and ALPI are found in many vertebrate species, whereas the two others are only found in mammals (5).

This study extends the knowledge of the mammalian TNSALP, which is the only ALP known to be involved in the human genetic disease, hypophosphatasia (OMIM 241510, 241500, and 146300 (6–11)). ALPL comprises 12 exons, of which the first and part of the second are noncoding. Notably, two transcripts have notably been identified, driven by two alternative exons 1 in the 5′-untranslated region (UTR) (6), which are responsible for differential transcription and subsequent post-translational modifications (12–14).

TNSALP is a homodimeric enzyme which each monomer is composed of 524 amino acids (57.2 kDa) (6, 12). TNSALP links to the cell membrane and functions as an ectophosphatase to hydrolyze monophosphated esters including inorganic pyrophosphate (PPi), phosphoethanolamine, and pyridoxal 5′-phosphate (15). Various domains were identified in TNSALP: (i) the active site, which includes a catalytic site and three ligands for metal ions (two zinc and one magnesium sites), and (ii) several mammalian-specific regions, i.e. the long N-terminal α-helix domain responsible for catalysis (16), the non-catalytic Ca2+-binding domain, and the crown domain (17). The crown domain provides several essential properties to the enzyme, namely, its competitive inhibition, its variable

* This work was supported in part by grants from the CNRS, Université Pierre & Marie Curie, and the patients’ support group Hypophosphatasie Europe.

† This article contains supplemental Tables S1–S3.

‡ Both authors contributed equally to this study.

§ Present address: Dept. of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel.

§ Supported by a “Société Française de Biologie des Tissus Minéralisés” student award.

¶ To whom correspondence should be addressed. Tel: 0033-144273572; Fax: 0033-144275801; E-mail: jean-yves.sire@upmc.fr.

5 The abbreviations used are: ALP, alkaline phosphatase; TNSALP, ALP tissue non-specific alkaline phosphatase; ALPI, placental alkaline phosphatase PLAP-1; GCAP, placental-like 2 or germ cell alkaline phosphatase; MEGAS, Molecular Evolutionary Genetics Analysis 5; HYPHY, Hypothesis testing using PHYlogenies; Ma, million years; SIFT, Sorting Intolerant From Tolerant.
heat stability, its allosteric behavior, and its interaction with the extracellular matrix, particularly collagen (18–20).

TNSALP seems to display post-translational modifications including five putative N-glycosylation sites (Asn\textsuperscript{140}, Asn\textsuperscript{230}, Asn\textsuperscript{271}, Asn\textsuperscript{303}, and Asn\textsuperscript{430}) and an undetermined O-glycosylation site (21, 22). Putative phosphorylation sites were not reported so far.

Hypophosphatasia is an autosomal disease caused by homozygous, heterozygous, or compound heterozygous ALPL mutations and can be either dominant or recessive. The phenotype is principally characterized by disorders in bone and tooth mineralization and clinical symptoms vary from moderate (as in the case of rickets, osteomalacia, or dental defects) to severe when deep hypomineralization of the skeleton occurs. Although this disease displays a low prevalence (1/300,000 in Europe), more than 275 mutations are reported in the literature to date (December 2013) (23) and of these 75% (204) result from amino acid substitutions (missense mutations).

The large range of genotypes and phenotypes associated with this disease makes it difficult to (i) establish genotype-phenotype relationships and (ii) predict the mutation severity because most of the patients have a unique genotype. Therefore developing tools for predicting and validating missense mutations leading to hypophosphatasia is crucial. Furthermore, it is critical that these tools are based on a deep understanding of the protein, particularly the importance of its functional and structural domains. Herein, we show that evolutionary analysis represents an extremely promising method for achieving this and we anticipate that it will have a significant impact in this field.

The increasing number of ALPL sequences found in public databases (NCBI and Ensembl) coming from genome and/or transcriptome sequencing in numerous vertebrate species enabled us to develop a new method for predicting disease mutations based on the evolutionary molecular analysis of protein genes (24, 25). This method involves comparing a large number of primary sequences of a given protein arranged into a phylogenetic context (i.e. based on species relationships) with the aim of identifying sensitive positions. Such evolutionary analysis allows us to (i) understand how the protein and its various regions evolved; (ii) identify regions, motifs, and residues that play important functions because they were conserved during long geological periods (hundreds of millions years: Ma); and (iii) predict and/or validate point mutations that could lead to a genetic disease, and ultimately estimate their degree of severity. Such a method, which uses evolutionary analysis to predict human diseases, was recently called phylomedicine (25). It was shown that at least 95% of the amino acid substitutions already known to be responsible for a genetic disease are validated with this method (24). Recently, our group successfully used this approach for predicting sensitive positions in various proteins of the secretory calcium–binding phosphoprotein family (26), and showed that missense mutations lead to genetic diseases. This approach was demonstrated on the following proteins: amelogenin (27), enamelin (28), matrix extracellular phospho-glycoprotein (29), amelotin (30), dentin matrix protein 1 (31), and ameloblastin.\footnote{F. Delsuc, B. Gasse, and J.-Y. Sire personal data.}

Here, we perform the evolutionary molecular analysis of mammalian TNSALP to (i) highlight the functional or structural importance of various positions and domains; (ii) predict sensitive positions that should be responsible for a genetic disorder when substituted; and (iii) validate our predictions by using missense mutations reported in humans.

**EXPERIMENTAL PROCEDURES**

**ALPL Sequences and Alignment**

A total of 58 ALPL sequences, representative of the main mammalian lineages, were extracted from NCBI and Ensembl databases. They comprised nine published, full-length cDNA sequences, 14 computer-predicted sequences available from sequenced mammalian genomes, and 35 sequences obtained using Basic Local Alignment Search Tool (BLAST) in genomes being currently sequenced (Table 1). The unpublished sequences were validated through alignment to the published cDNA sequences using Se-Al v.2.0a11 software (32). The intron-exon boundaries were carefully checked. When necessary, the sequences were completed and/or corrected either using Blast search in the Ensembl genome database, in the NCBI trace archives, and in the whole genome shotgun repository sequences, or using de novo sequencing using PCR on genomic DNA extracted from ethanol-preserved tissues (see Table 2 for primers used). The 58 nucleotides sequences were translated into amino acid sequences and aligned to human TNSALP chosen as reference sequence in our alignment. The 58 coding ALPL sequences are available online.

Our final alignment consisted of 528 positions, including only four insertions (supplemental Table S1). A few residues were missing in this alignment (395 nucleotides), representing less than 1.3% of the data, and the corresponding positions were treated as “unknown data.”

**Evolutionary Analyses**

**Substitution Models**—The substitution models for both the nucleotide and amino acid alignments were defined using an online model selection tool datamonkey (33). Briefly, for each alignment, the software defines the best model by testing all possible evolutionary models. Here, the two best substitution models were (i) the Tamura-Nei model (34) for the nucleotide alignment; and (ii) the Jones et al. model (35) for amino acid alignment.

**Distance Tree**—This approach is useful for the evolutionary analysis as it highlights variable sequences related to either large evolutionary distances or high substitution rates with regard to amino acid conservation through mammalian evolution. The 58 aligned mammalian ALPL sequences were computed using Molecular Evolutionary Genetics Analysis 5.2.1 (MEGA5) (36). The alignment was restricted to the length of the human ALPL sequence by excluding the 12-nucleotide insertions in non-human sequences and the stop codon. This procedure resulted in a dataset containing 1572 nucleotide sites.
**Evolutionary Analysis of Sensitive Positions in TNSALP**

**TABLE 1**

Preferred common names, scientific names, families, orders, and references in GenBank™ for the 58 mammalian species, from which ALPL sequences were used in our study.

The species are arranged in alphabetical order of common names.

| Common name      | Genus and species | Family                        | Order      | Source               |
|------------------|-------------------|-------------------------------|------------|----------------------|
| Aardvark         | Orycteropus afer  | Orycteropodidae               |            |                      |
| Armadillo        | Dasypus novemcinctus | Dasypodidae               |            |                      |
| Aye-Aye          | Daubentonia madagascariensis | Daubentoniidae |            |                      |
| Baboon           | Papio anubis      | Cercopithecidae               |            |                      |
| Big brown bat    | Pteropus vampyrus | Pteropodidae                  |            |                      |
| Babiroussas      | Galagidae         | Scandentia                    |            |                      |
| Camel            | Camelus ferox     | Camelidae                     |            |                      |
| Cat              | Felis catus       | Felidae                       |            |                      |
| Chimpanzee       | Pan troglodytes   | Hominidae                     |            |                      |
| Chinchilla       | Chinchilla laniger | Chinchillidae                |            |                      |
| Chinese hamster  | Cricetulus griseus | Cricetidae                  |            |                      |
| Cow              | Bos taurus        | Bosidae                       |            |                      |
| Degu             | Octodon degus     | Octodontidae                  |            |                      |
| Tenrec           | Echinosorex jilaxi | Ctenodactylidae              |            |                      |
| Dolphin          | Tursiops truncatus | Delphinidae                  |            |                      |
| Elephant         | Loxodonta africana | Elephantidae                 |            |                      |
| Elephant shrew   | Elephas maximus    | Proboscidea                   |            |                      |
| Eretabat         | Mustela putorius furo | Mustelidae                |            |                      |
| Gibbon           | Nomascus leucogenys | Primiatae                   |            |                      |
| Goat             | Capra hircus      | Bovidae                       |            |                      |
| Golden mole      | Chrysoschis asiatica | Chiroptera                  |            |                      |
| Gorilla          | Gorilla gorilla   | Primates                      |            |                      |
| Guinea pig       | Cavia porcellus   | Rodentia                      |            |                      |
| Hedgehog         | Erinaceus europaeus | Chiroptera                  |            |                      |
| Horse            | Equus caballus    | Equidae                       |            |                      |
| Human            | Homo sapiens      | Chiroptera                    |            |                      |
| Hyrax            | Procavia capensis | Rodentia                      |            |                      |
| Jerboa           | Jaculus jaculus   | Carnidae                      |            |                      |
| Kangaroo rat     | Dipodomys ordii   | Rodentia                      |            |                      |
| Killer whale     | Oryctolagus Jaculus | Hominidae                   |            |                      |
| Macaque          | Macaca mulatta    | Cercopithecidae               |            |                      |
| Manatee          | Trichebus manatus | Rodentia                      |            |                      |
| Marmoset         | Callithrix jacchus | Cebidae                      |            |                      |
| Megabat          | Pteropus vamyrus | Cebidae                       |            |                      |
| Microbat         | Myotis lucifugus  | Cebidae                       |            |                      |
| Mouse lemur      | Microcebus marinus | Cricetidae                  |            |                      |
| Mouse            | Mus musculus      | Rodentia                      |            |                      |
| Naked mole rat   | Heterocephalus glaber | Cricetidae                |            |                      |
| Opossum          | Monodelphis domestica | Rodentia                   |            |                      |
| Orangutan        | Pongo pygmaeus    | Rodentia                      |            |                      |
| Panda            | Ailuroprocta melanoeuca | Ursidae                   |            |                      |
| Pig              | Sus scrofa        | Rodentia                      |            |                      |
| Pika             | Ochotona princeps | Mustelidae                    |            |                      |
| Platypus          | Ornithorhynchus anatinus | Rodentia               |            |                      |
| Prairie vole      | Microtus ochrogaster | Cricetidae                |            |                      |
| Rabbit           | Oryctolagus conulosus | Rodentia                 |            |                      |
| Rat              | Rattus norvegicus | Rodentia                      |            |                      |
| Rhinoceros       | Ceratotherium simum | Rodentia                   |            |                      |
| Sheep            | Ovis aries        | Rodentia                      |            |                      |
| Shrew            | Sorex araneus     | Rodentia                      |            |                      |
| Squirrel         | Idioryctina tridecareus | Rodentia               |            |                      |
| Squirrel monkey  | Samiri boliviensis | Rodentia                    |            |                      |
| Star-nosed mole  | Condylura cristata | Rodentia                   |            |                      |
| Tasmanian devil  | Sarcophilus harrisii | Rodentia                  |            |                      |
| Tree shrew       | Tupaia chinensis  | Rodentia                      |            |                      |
| Wallaby          | Macropus eugenii  | Rodentia                      |            |                      |
| Walrus           | Odobenus rosmarus | Rodentia                      |            |                      |

**TABLE 2**

Primer sets designed to amplify ALPL exons from genomic DNA in five mammalian species to either complete or correct sequences partially found in databases.

Expected PCR product sizes (base pairs) and PCR annealing temperatures (°C) are indicated. See Table 1 for Latin names.

| Common name      | Targeted exon | Forward (5' → 3') | Reverse (5' → 3') | Base pairs | °C   |
|------------------|---------------|-------------------|-------------------|------------|-----|
| Dolphin          | 10            | TAAACCCTTGCTTGCAGTTC | ACCCGAGCTACCTCTACCT | 580        | 59  |
| Hedgehog         | 4             | ACCGACTCGAGGCAGTACA | TCTTTGACTCTCGGTCCACC | 444        | 59  |
| Hedgehog         | 6             | GGGTTCTCCTGCTGCTCTGGA | CAGGCCTGCTGCTACCTG | 401        | 59  |
| Mouse lemur      | 9             | AAGAAGCAGCGGCCAGACA | CCTGACAGCTACGCTCTC | 771        | 56  |
| Mouse lemur      | 12            | AGGTGGGCACTGGCTGCCTCT | AATCCGAGGAGGGCAGAT | 598        | 59  |
| Panda            | 5             | GAAATCCATCCCCGCTTGGA | AGTGCGACAGGCTCTCACCT | 442        | 62  |
| Panda            | 12            | CTCGCGTACGAGGAGCAGT | GGTGCGTACGCTGCTCTTTC | 599        | 59  |
| Tree shrew       | 8             | ACTGCTCTTGTTGCTACTTCT | AGGAGAATTTGCTCTGTGCT | 376        | 59  |
TABLE 3
Favored, neutral, and disfavored amino acid substitutions in proteins

The order of amino acids shows the substitution preferences, from the favored substitutions to the most disfavored. From Ref. 41.

| Residues | Favored | Neutral | Disfavored |
|----------|---------|---------|-----------|
| Ala (A)  | R, S, T, E, V, G, H, I, K, M, L, N, P, Q | F, Y, D, W, C | A, Y, V, T, W, F, H, L, L, M, Q, R, S, G, E, P, K, N, D |
| Cys (C)  | N       | T, E, K, G, H, P, Q, R, S | A, V, M, L, F, L, Y, W, C |
| Glu (E)  | Y, W    | Q, P, D, N, K, H, M, R, S, T, V, A | L, G, W, V, Y, F, C |
| Gly (G)  | A, Q, P, D, T, N, H, R, S | L, V, I, M | A, R, H, T, K, D, E, N, Q, P, S, G, C |
| His (H)  | A, P, N, D, E, G, K, Q, R, S, T, Y | A, V, M, L, V, W, Y, F, C |
| Ile (I)  | V, L    | A, K, M, R, T, F, Y | N, H, W, E, P, Q, S, D, G, C |
| Lys (K)  | R       | Q, P, D, E, N, H, I, A, S, T, V | M, L, G, Y, F, W, C |
| Met (M)  | N       | A, P, F, T | S, Y, E, H, K, Q, R, N, G, D, W, C |
| Pro (P)  | L       | A, V, T, R, E, F, Q, I | H, K, N, P, S, W, Y, G, D, C |
| Glu (Q)  | D       | A, E, T, G, H, K, P, Q, R, S | M, I, Y, V, F, L, W, C |
| Arg (R)  | K       | N, D, E, M, G, H, I, A, Q, P, S, T, V, Y | L, W, Y, I, F, C |
| Ser (S)  | T       | P, D, E, N, H, G, K, R, A, Q | F, W, L, C |
| Thr (T)  | S       | A, V, D, E, G, H, I, K, M, L, N, P, Q | I, M, L, V, W, Y, F, C |
| Val (V)  | L       | A, Q, R, T, E, F, Y, K, M, P | W, F, Y, C |
| Trp (W)  | F, Y    | N, H, D, W, S, G, C | N, H, D, W, S, G, C |
| Tyr (Y)  | F, W    | V, I, H, R | R, Q, E, M, H, T, V, L, S, A, G, K, L, N, D, P, C |
|          |         |         | A, E, K, M, L, N, P, Q, S, T, G, D, C |

(524 amino acids). This alignment was supplemented by a mammalian tree taking into account the recent mammalian phylogeny (37).

Maximum likelihood phylogenetic reconstruction was performed from the nucleotide alignment using the Hypothesis testing using Phylogenies (HYPHY) software (38). For each nucleotide, the probability for the observed data was calculated using the maximum likelihood method and taking into account the supplied tree.

Analysis of Functional Constraints—As previously described (29–31), the strong constraints corresponding to structural/functional regions were searched on the nucleotide and amino acid alignments by Sliding Windows analysis in the HYPHY software and by SLAC analysis (datamonkey) (39, 40). For the Sliding Windows analysis, the logarithm of the probability was calculated for a window of 15 base pairs (bp) with an overlap of 5 bp between each window and using a likelihood algorithm taking into account phylogenetic relationships.

Analysis of Amino Acid Substitutions—Site-specific selections (i.e. biologically significant amino acids) in TNSALP were identified in our alignment for the 524 positions and displayed on the human sequence. The analysis was performed according to the substitution preferences of amino acids, i.e. favoring property conservation (41) (Table 3). We defined three levels of selection: conserved (i.e. unchanged residues during 220 Ma), conservative (i.e. substituted residues having similar properties), and variable positions (i.e. substitution with various residues).

Prediction of Post-translational Modifications—We looked for putative N-glycosylation and phosphorylation sites (casein kinase II, protein kinase C, and tyrosine kinase phosphorylation sites) in human and some mammalian sequences using MyHit database). We calculated the theoretical pI (isoelectric point) of different motifs with the Compute pI/Mw tool (Expasy).

RESULTS

Sequence Comparison of Mammalian TNSALP

With a total of 58 species from 22 orders, our dataset is well representative of the current mammalian TNSALP diversity and it covers circa 220 Ma of evolution (Table 1). The length of the coding ALPL sequence, from ATG to the stop codon, was found identical in 53 species: 1575 bp resulting in 524 amino acids from the N-terminal methionine, Met1, found in the region encoded by exon 2, to the phenylalanine preceding the stop codon, Phe224, encoded by exon 12 (supplemental Table S1). Short variations were only observed in five TNSALP.

In the region encoded by exon 2, one or two amino acids were deleted before the cleavage site of the signal peptide in Marsupialia (wallaby, Tasmanian devil, and opossum); encoded by exon 3, two deletions and one insertion are found in an Afrotheria (hyrax); encoded by exon 11, one deletion in a Chiroptera (megabat); and encoded by exon 12, one deletion in opossum and wallaby, and three insertions in Marsupialia. Aside these variations, 303 residues were found unchanged in TNSALP during 220 Ma of evolution (supplemental Table S1).

Distance Tree

To define whether all sequences were relevant for our evolutionary analysis, we calculated the pairwise distance for each sequence in the alignment. This method allowed to quantify the evolutionary distance for each taxon. This distance was shown on a maximum likelihood tree (Fig. 1). First of all, we observed that the substitution rates were low for TNSALP (0.11) compared with unordered proteins previously studied in our group (DMP1, AMTN, and MEPE), which display larger branch lengths indicating rapidly diverging sequences (for instance, 0.332 and 0.602 for hedgehog and platypus DMP1, respectively). Substitution rates of TNSALP were particularly low in primates, as illustrated by short branches, e.g. from 0.00365 between human and chimpanzee to 0.01199 between marmoset and squirrel monkey, except for bushbaby (0.07601). Other sequences have higher substitution rates as shown by the long branches of platypus (0.11082), opossum (0.08267), Tasmanian devil (0.07516), and shrew (0.06991), but taken together the low substitution rates demonstrated that all sequences were relevant for the evolutionary analysis and that TNSALP was subjected to a high selection pressure in all species.
Analysis of Functional Constraints

Sliding Window analysis indicated that the TNSALP sequence is characterized by an alternance of high and low values of maximum likelihood, reflecting weak and strong selective pressures, respectively (Fig. 2A). The non-synonymous substitution rate (dN) obtained with SLAC analysis confirmed these observations and allowed to identify various domains more precisely (Fig. 2B). The region encoded by the 3’ extremity of exon 2, corresponding to the signal peptide cleavage site, is well conserved, indicating a strong functional pressure. The regions encoded by the 3’ and 5’ extremities of consecutive exons are also strongly constrained, especially exons 3–4, 4–5, 9–10, and 11–12. The analysis confirms that several, already known, functional motifs are subjected to selective pressure: the homodimeric interfaces principally those encoded by exons 4 and 5, and the active sites encoded by exons 5, 6, 10, and 12.

Putative Post-translational Modifications

In human TNSALP MyHit confirmed the presence of the five putative N-glycosylation sites (motif = NX(S/T)/C) already reported in the literature (140NX142, 230NX132, 271NX273, 303NX1305, and 430NX1432) (Fig. 3). Although these sites are generally well conserved through 220 Ma, our evolutionary analysis revealed that some mammalian TNSALP are lacking correct components of the site: 140NX142 was changed into 140TX142 in pika and into 140SX142 in panda; 230NX132 was replaced with 230XP132 in opossum; 271NX273 was replaced with 271NXM273 in opossum and wallaby; and 430NX1432 was 430NX1432 in panda and 430NXF1432 in Tasmanian devil. Only 303NX1305 was kept functional during mammalian evolution, although slightly changed in cat: 303NX1304 (see supplemental Tables S1 and S2).

In addition to N-glycosylation sites, 18 putative phosphorylation sites were identified: nine for casein kinase II ((S/T)XX(E/D)): 85TLR88, 188DSNE191, 232TDVE235, 245TLR248, 308SLSE311, 366TSSE369, 372SDSE375; seven for protein kinase C ((S/T)XX(R/K)): 36TLK38, 134TER136, 166TTR168, 255TWK257, 258SFK260, 389TPR391, and 405TDK407; and two for tyrosine kinase ((R/K)XX(D/E)XXY): 22KEKDPKY28 and 229KNKTDVEY236 (Fig. 3). Our analysis indicated that five were conserved during mammalian evolution. They are: three casein kinase II (188DSNE191, 245TLR248, and 308SLSE311) and two protein kinase C (166TTR168 and 389TPR391) phosphorylation sites. MyHit did not identify any putative O-glycosylation site.

Analysis of Amino Acid Substitutions

Of the 524 amino acids composing the human TNSALP sequence, 469 (89.5%) were identified as sensitive positions, i.e. that were either conserved (i.e. unchanged residues) (303 residues, 57.8%) or conservative (i.e. substituted with residues having similar properties) (166 residues, 31.7%) during 220 Ma of mammalian evolution. Only 55 positions (10.5%) were identified as variable (i.e. substituted with various residues). These unchanged, conservative, and variable positions were reported on the human sequence, which constitutes the chart of sensitive positions of human TNSALP (Fig. 3). We predict that any substitution occurring at one of the 303 unchanged positions or any substitution of one of the 166 conservative positions with a residue having a different property would disrupt TNSALP function and lead to hypophosphatasia.

All substitutions on conservative and variable positions found in the 58 TNSALP sequences were reported in the Online Resource 4. Our analysis confirms that the protein is strongly constrained and indicates that each of the 469 sensitive positions plays an important role of either structural or functional significance.

It is worth noting that the percentage of purifying selection is high along the protein sequence, and varies from 80% in the region encoded by exon 2 to 96.6% in the region encoded by exon 6. More precisely, our evolutionary analysis (i) confirmed already known important sites and domains of the protein and (ii) highlighted new sites having putative important roles (Fig. 3).
Confirmation of Important Sites and Domains

The residues and motifs known to play an important role for the right function of TNSALP are briefly reviewed below, along the protein sequence, at the light of our evolutionary analysis. See “Discussion” for a more accurate description of the conservation of these important positions.

The region encoded by exon 2 (80% of purifying selection): 4 unchanged positions, signal peptide and cleavage site, Ser^{17}, conserved; exon 3 (82.5%), 20 unchanged positions; homodimeric interface (region from Lys^{27} to Glu^{62} housing seven unchanged and three conservative positions), N-terminal helix, and active site, Asp^{60}, conserved; exon 4 (90%), 28 unchanged positions, homodimeric interfaces conserved (three domains), from Thr^{68} to Gln^{76} (eight unchanged and one conservative), from ^{80}EET^{85} (one unchanged, one conservative, one variable), and from Phe^{94} to Ser^{98} (four unchanged, one variable); exon 5 (94.8%), 43 unchanged positions, homodimeric interfaces (one motif, ^{103}TNA^{105}, and one residue, Gly^{120}, unchanged) and active site, Asp^{109}-Ala^{111}, conserved; exon 6 (96.6%), 37 unchanged positions, active site, His^{171}, Thr^{173}, Arg^{184}, and Tyr^{187} conserved; exon 7 (87.5%), 23 unchanged positions; exon 8 (87%), 9 unchanged positions; exon 9 (91.1%), 27 unchanged positions, active site, Glu^{132}, conserved; exon 10 (89%), 46 unchanged positions, homodimeric interface (from Val^{382} to Arg^{391}, with eight unchanged and two conservative residues), and active site Asp^{537}, His^{541}, Arg^{578}, His^{579}, and His^{581}, conserved; exon 11 (92.5%), 25 unchanged positions, homodimeric interface conserved (from Gly^{597} to Phe^{610} with eight conserved, five conservative and one variable positions); exon 12 (85.2%), 41 unchanged positions, homodimeric interface (from Ser^{645} and Ile^{690}, with 28 unchanged residues and 18 conservative positions) and active site, His^{654}, conserved.

In addition to these conserved positions, two previously identified large regions encoded by several exons, the Ca^{2+}-binding domain and the crown domain, are also well conserved (Fig. 3).

Other Sites

In addition to the important regions confirmed above, various residues of unknown function are well conserved in numerous regions, and are located far from the active sites. Three motifs seem to play an important role: ^{126}GTVGVSAA^{133} and ^{144}GNEVTSILRWA^{154}, encoded by exon 5, and ^{348}ALHEAVEMDRAI^{359}, encoded by exon 10. Three-dimensional modeling shows that these three regions are located at the surface of the molecule (Fig. 4: see “Discussion”).

Validation of ALPL Mutations in Humans

Our Predictive Evolutionary Chart—So far, of the 204 missense mutations listed in the ALPL mutation database, 145 positions only are affected as more than one amino acid substitution can occur in a same position (supplemental Table S2). These mutated positions were indicated on the human TNSALP sequence and compared with the sensitive positions identified in our evolutionary analysis (Fig. 3). Among the 145 positions affected in humans, 113 concerned unchanged positions and 31 conservative positions were identified as sensitive positions in our evolutionary analysis, and one variable position. All human mutations on a conserved position and 8 on a conservative position were directly validated by our evolutionary analysis, which represents 82.8% of the affected positions in humans. In 23 of 24 conservative positions, the substituting amino acids were never found in the nature (supplemental Table S2). Although our evolutionary analysis could not predict initially such substitutions as being deleterious, the fact that
they were not shared by any mammalian TNSALP validates them secondarily. The same reasoning can be applied to the only variable position Asp406, in which two deleterious substitutions were found in humans (p.D406N and p.D406G).

Indeed, the two residues substituted were not found in the wild (supplemental Table S2). Therefore, our evolutionary analysis validated 99.3% of the deleterious missense mutations observed in human TNSALP.

Eventually, only one missense mutation (0.7%) did not fit with our evolutionary analysis. This mutation was reported on a conservative position (p.A177T), but was not predicted as deleterious because the same substitution occurred in the wild (supplemental Table S2).

Other in Silico Prediction Tools—To date, several in silico models dedicated for predicting the effects of amino acid substitutions in humans are available and easy to access (for review, see Ref. 42). We list the results obtained with TNSALP using five softwares (see Table 4 and supplemental Table S3). To predict the effect of the substitutions, there are three categories of tools: the sequence and evolutionary-conservation-based methods, such as SIFT or Mutation Assessor, which base their prediction on sequence conservation using a multiple sequence alignment; the protein sequence and structure-based methods, such as PolyPhen2 or MAPP, which combine amino acid properties, location of functional site, secondary structure, and membrane topology; the supervised-learning methods like SNAP, which include for their algorithms neural networks, support vector machines and random forests, and naive Bayes classifiers. The predicted mutations are considered as true positives, the nonpredicted mutations as false negatives, and the false predictions as false positives.

PolyPhen-2 is software based on both sequence and structural information, and classifying mutations as deleterious or benign. This tool provided a correct prediction, for 93.6% of the missense mutations known in ALPL, but erroneously considered 47 substituted positions, i.e. 9% of the 524 residues, with amino acids found in the nature as damaging the protein function.

MAPP (Multivariate Analysis of Protein Polymorphism) is software combining both evolution and physicochemical characteristics like hydropathy, polarity, charge, volume, free energy in α-helix, and β-strand conformations. This powerful software detects “good” and “bad” amino acids for each position. All known mutations of TNSALP were found except K264R and R450H. Nevertheless, MAPP also detected as bad 139 substituted positions (26.5%) with residues found in nature.

Mutation Assessor is designed to predict the functional impact of amino acid substitutions, through evolutionary conservation in the protein family and subfamily. This software provided a correct prediction for 82.8% of TNSALP missense mutations. A total of 109 moderate phenotypes were predicted by the software, but 10 of the 24 moderate phenotypes reported in the literature were not validated. Similarly, Mutation Assessor predicted 59 positions as leading to a severe phenotype when substituted, but did not validate seven of the severe phenotypes known in the literature. Also, Mutation Assessor wrongly detected 48 substituted positions (9.1%) with residues found in nature as leading to diseases.

SIFT (Sorting Intolerant From Tolerant) identifies deleterious mutations using only an evolutionary approach. This method is based on the degree of conservation of amino acids in sequence alignments derived from closely related sequences obtained by PSI-BLAST. This software did not detect 50% of
the mutations known in the literature and erroneously predicted 30 amino acids (5.7%) as nontolerated although they are found in nature (e.g. K27W, N53K, ...).

SNAP (Screening for Non-acceptable Polymorphisms) is a neural network using derived protein information as, for instance, secondary structure, evolutionary conservation, and solvent accessibility to predict the effect of mutations on the protein. This tool validated 69% of the missense mutations reported in the literature for TNSALP, but performed false prediction for 18 positions (3.4%) in considering residues as non-neutral for the protein function, whereas there are observed in nature. These predictions and validations are commented upon under “Discussion.”

**DISCUSSION**

The evolutionary molecular analysis of mammalian TNSALP yields important new information by (i) pointing precisely to residues and motifs of crucial structural and/or functional importance through evidence of selective constraints; (ii) predicting which positions could result in a genetic disorder when substituted; and (iii) validating all but one missense mutation reported so far in humans. In addition, in conservative and variable positions, confronting deleterious missense mutations in humans and in other mammals allowed to detail the range of possible substitutions at these positions in TNSALP. Selective Constraints and Evidence of Functional Residues and Motifs—First, our study demonstrated that TNSALP is a highly constrained protein as illustrated by 89.5% sensitive positions, identified through 220 Ma of mammalian evolution. This high percentage of conservation during such a long geological period reflects strong selective constraints acting on the protein at the amino acid level. This also reveals that the protein needs to be highly structured on its full-length for a correct function of the enzyme. This high selective pressure along the whole sequence length contrasts with the low constraints observed in unstructured (disordered) proteins previously studied: AMELX (40% of sensitive positions (27)), ENAM (7% (28)), MEPE (14% (29)), AMTN (8% (30)), DMP1 (20% (31)), and AMBN (18%).7 This corroborates the crucial importance of TNSALP in mammals and its involvement in various functional, but also structural, roles, not yet exhaustively identified.

Active Sites and Their Neighboring Residues—We found that the 14 previously identified active sites of TNSALP and their flanking residues were unchanged during 220 Ma, which confirms their crucial importance for the right function of the protein. More precisely, the 58LGDGM62 motif, including Asp60, which plays a role as ligand for metal ions, Zn2⁺ (zinc site 2) and Mg2⁺ (43); 107VPDSAG122 motif, with Ser110, which covalently binds phosphate during catalysis (44). The residue Tyr117, located close to this active site, is highly conserved. This amino acid is located in the 12-Å sphere surrounding the phosphate group; the 171HATPS175 motif, containing His171 and Thr173, described as indirect ligand and ligand of Mg2⁺ (magnesium site), respectively; 332EGGIRDGHHH341 motif, including two residues of the magnesium site, Glu332, which is a direct ligand, and His338, an indirect ligand, and two ligands of Zn2⁺ (zinc site 1), Asp337 and His341. In addition, in this motif, His340 is involved in a hydrogen bond and plays an important role in stabilizing the active site environment (45); the 377ADHSH381 motif, in which Asp378, His379, and His381 are occupied by a Zn2⁺ (zinc site 2); and finally, the 451HETHG455 motif, housing His454, which is important for the zinc site 2. TNSALP also contains a Ca2⁺-binding site, which is non-catalytic and possesses 76 residues (from Gly222 to Leu309), including Phe290, Glu291, and Asp307 that were unchanged during evolution and bind calcium (17, 46).

Peripheral Binding Site—The Ca2⁺-binding site also includes a peripheral binding site, which is constituted by residues located on two α-helices: Trp270, Leu276 and Ser308, Glu318 (47). Among them five residues (Trp270, Arg272, Leu275, Asp306, and Glu311) were unchanged during mammalian evolution. This finding highlights their crucial importance for the structure/function of the enzyme, although their precise role is still unknown.

Disulfide Bonds—The two disulfide bonds of TNSALP, Cys139-Cys201, and Cys489-Cys497, were unchanged, indicating their important role in the stability of the protein structure (45).

N-terminal Fragment—The domain Lys27-Glu42 located in the N-terminal region forms an α-helix, which interacts with specific residues to stabilize the protein and allow catalysis (16). Of the 16 residues composing this domain, seven were unchanged: Tyr28, Trp29, Glu32, Glu34, Thr36, Leu37, and Leu41. This finding underlines the essential role that these amino acids play for TNSALP function and/or structure.

Crown Domain—The crown domain is an essential domain, exclusively found in mammalian TNSALP. This domain is composed of 65 amino acids (from Gly387 to His451), most of them being well conserved during evolution (only three variable positions), a finding that points to the important function of this domain. It includes two residues essential for their competitive inhibition properties: Tyr388, which is unchanged, and His451, which is on a conservative position. The latter residue plays a role in substrate binding of the nearby active site (see above). Within this crown domain, the region Gly425-His451 was shown to be a collagen binding site (20). Our evolutionary analysis confirmed the importance of this region and allowed to measure the precise length of this binding site in shortening the functional domain to 16 amino acids, Val434-Leu449. This motif is more acidic, with a pI of 5.08, allowing the electrostatic interaction with the N-telopeptide of the collagen fibrils. However, the high affinity of this region has to be tested in vitro.

Ionic Pocket—A highly ionic pocket was described as formed by six residues (Asp109, Glu125, Gly126, Arg184, Tyr187, and Tyr388) (43). These positions are unchanged excepted for Glu125, which is substituted with His125 in Tasmanian devil (Marsupialia). The ionic pocket is specific to TNSALP and these residues are described as constituting an hydrophobic pocket in other ALPs.

Homodimeric Interface—Le Du et al. (17) identified 83 amino acids involved in the homodimeric interface. Using the three-dimensional model published by Mornet et al. (46) and our evolutionary analysis, we identified nine residues or motifs playing an important role in this interface. Indeed, of the 83

7 F. Delsuc, B. Gasse, and J.-Y. Sire, personal communication.
Evolutionary Analysis of Sensitive Positions in TNSALP

amino acids involved in the homodimeric interface, 92% are on sensitive positions (64% unchanged and 28% conservative), which allowed us to conclude that the homodimeric interface is highly constrained and plays an essential role for the correct function of the protein. This must be related to the observation that ALPs are active only in homodimeric form in all species, including prokaryotes (48).

Post-translational Modifications—Although the five putative N-glycosylation sites identified in human TNSALP were rather well conserved during mammalian evolution, slight differences observed in some species indicate that four of these sites were probably not functional in the last common mammalian ancestor; only 300NXT-305 was kept unchanged during mammalian evolution. We can conclude that either the four other sites are not glycosylated or that glycosylation occurred late in mammalian evolution: in this case, 230NXP232 and 271NXT-273 glycosylations could have occurred in an ancestral Placentalia, and 140NXT142 and 430NXS432 in an ancestral Primate. Nosjean et al. (22) showed that cleavage of N-glycosylations removed human TNSALP activity, with the exception of Asn230 and Asn303. This finding seems to support that Asn140, Asn271, and Asn430 glycosylations occurred recently in mammalian evolution and that Asn230 is not glycosylated. In contrast the evolutionary conservation of Asn303 could indicate either the presence of a N-glycosylation or another function for this site. Studying TNSALP in nonmammalian species could bring more accurate information on the story of these sites.

This study reports for the first time the presence of putative phosphorylation sites. Of 18 sites, five were found unchanged during mammalian evolution and could be functional, but their presence should be experimentally tested.

New Putatively Important Sites—Three highly conserved motifs were not currently reported in the literature as being functionally or structurally important for TNSALP (Fig. 4). 126GTVGVSA133 (Fig. 4A) and 144GNEVTSILRWA154 (Fig. 4C) are both encoded by exon 5 (chain X), but in the first motif, the first three amino acids are located in the 12-Å sphere around the phosphate group (49). The third motif, 348ALHEAVEMDRAI359 (Fig. 4B), is encoded by exon 10 (chain Y) and contains an acidic cluster, giving a pl of 4.65, more acidic than the collagen binding site. These three sequences are not referenced in the peptide database. Interestingly the three-dimensional model shows that these three motifs are at the surface of the molecule and very exposed, A and C on the front side and close enough to each other to form a unique domain, and B on the rear side (Fig. 4). A and C motifs are also very close to the N-terminal α helix of the other monomer and probably contributes to the homodimer interface, whereas the role of motif B remains to be elucidated.

Predictions and Validations of Human Mutations—Our evolutionary molecular analysis of mammalian TNSALP highlights the functional and/or structural importance of 469 unchanged or conservative positions on the human sequence. All these positions are predicted to be sensitive, i.e. their change will result in hypophosphatasia. By contrast, three well known polymorphisms of the ALPL gene, p.R152H, p.Y263H, and p.V522A, were found in conservative (p.R152H) and variable (p.Y263H and p.V522A) positions, and the conservative histidine residue at position 152 was observed in various other species, supporting the benign effect of the change. Thus, this predictive chart of sensitive positions on the human TNSALP sequence could be a useful tool for clinicians when identifying a new missense mutation of ALPL in a patient.

When comparing the 145 positions known to be affected in human TNSALP (missense mutations) to our predictive chart (Fig. 3), 120 of these positions (82.75%) correspond to mutations that would have been directly predicted to be deleterious because they occurred on sensitive positions. Of the unpredicted 25 missense mutations, 23 are located on a conservative position that was substituted with a residue having similar chemical properties but was never found in nature. This means that (i) our table of residue substitution at each position in the 58 mammalian species (supplemental Table S2) validates the deleterious effect of substitutions observed in humans, increasing to 98.6% the percentage of prediction, and (ii) the tolerance for changes affecting conservative positions must be more accurately specified in terms of similar properties. Indeed, all residues sharing similar chemical properties cannot substitute the functional amino acid. Structural and/or three-dimensional properties should also be considered probably in relationship with the highly constrained structure of the protein.

The two last missense mutations of TNSALP that were not directly predicted by our chart concern (i) conservative position 177 that is similarly mutated in nature and (ii) the variable position 406. By definition, a variable position could be substituted by any residue without any deleterious effect. This is not the case for this variable position. In humans, two missense mutations are reported on this position, p.D406N and p.D406G, but these substitutions were never observed in the 58 species studied that are well representative of the mammalian lineages. In other words, if the aspartic acid at position 406 could be substituted by either asparagine or glycine without any deleterious effect, such a substitution would have occur at random in one of the studied species during 220 Ma of evolution. Here again our substitution table predicts deleterious substitutions, probably related to large differences in the properties of the substituted residues, and increases to 99.3% the percentage of prediction. The human mutation p.A177T on a conservative position was unpredictable as the substituting amino acid reported responsible for the disease in human (Thr) was also found in three mammalian species (Myotis lucifugus, Chrysoschoris asiatica, and Echinops telfairi). The mutation was reported by Goseki-Sone et al. (50) in a patient with adult hypophosphatasia and the authors tested 111 unrelated genomic DNAs to exclude a polymorphism. It remains possible, however, that this mutation is in fact a very rare polymorphism in the human population.

Among the 204 missense mutations reported so far in the database of ALPL mutations responsible for hypophosphatasia, the clinical spectrum of patients is highly variable, ranging from the most severe form, presenting with stillbirth without mineralized bone, to the most moderate form presenting with loss of tooth without bone symptoms. On the basis of clinical reports, recurrent genotypes, homozygous genotypes, and site-directed mutagenesis experiments performed in our group or by others (see references in the ALPL mutation database) we were able to
Evolutionary Analysis of Sensitive Positions in TNSALP

| TABLE 5 | Severity of ALPL gene mutations versus conservation of residues |
|-----------------|------------------|------------------|------------------|
|                | Conserved position | Conservative position | Variable position | Total |
| Severe alleles  | 140 (87.5%)       | 19 (11.9%)         | 1 (0.7%)         | 160   |
| Moderate alleles| 15 (62.5%)        | 9 (37.5%)          | 0                | 24    |
| Total           | 155               | 28                | 1                | 184   |

qualify 184 mutations as severe or moderate alleles. Eighty-seven percent (140 of 160) of severe alleles were at conserved positions, 12% (19/160) at conservative positions, and one at a variable position (p.D406N), whereas 37% (9/24) of moderate alleles were found at conservative positions and 63% (15/24) at conserved positions (Table 5). Thus, although the evolutionary chart is efficient to distinguish polymorphisms and mutations, it remains relatively ineffective at predicting the severity of mutations. However, we can observe that five positions are subjected to various mutations and the functional consequences of such mutations are different: A116S has a moderate phenotype, whereas A116T has a severe one; R136C has a severe phenotype, whereas R136H and R136L have a moderate one; F327C and F327G have a severe phenotype, whereas F327L has a moderate phenotype; I395T has a severe phenotype, whereas I395V has a moderate phenotype; and R450C has a severe phenotype, whereas R450H has a moderate phenotype. According to the substitution preferences of amino acids, i.e. favoring property conservation (41) (Table 3), we observed that the most favored substitution leads to a moderate phenotype, whereas the most disfavored substitution leads to a severe phenotype. These findings confirm that the nature of the amino acid change plays a determining role in the final function of the enzyme at both conservative and conserved positions. These observations also suggest that it may be interesting to combine such sequence alignment results with algorithms predicting the effect of the change in terms of secondary structure, spatial constraint, polarity, and biochemical properties.

The five in silico models dedicated for predicting the effects of amino acid substitutions in humans were not as efficient predictive tools as our evolutionary chart, even if these predictive tools that try and use structural factors for predicting deleterious mutations are at a disadvantage because there are many ways a protein can adapt to a site mutation that cannot be known a priori (Table 4 and supplemental Table S3). For example, there are the compensated pathogenic deviations that are compensatory mutations found in a protein of one species cancelling the effect of disease-associated mutations in the ortholog protein in another species (1, 2). These adaptations can explain why we observed seven residues (N53K, V67R, D233H, V331S, T384P, D406Y, and 465S) predicted as deleterious by the five in silico tools even if they are observed in other mammalian genomes. PolyPhen-2 and MAPP provided accurate prediction of sensitive positions in the TNSALP sequence (93.6 and 99%, respectively), but a number of amino acids were erroneously predicted as deleterious (9 and 26.5%, respectively). Indeed, we can deduce that the substitution is neutral for these residues being observed in nature. Mutation Assessor, SIFT, and SNAP were less efficient in providing accurate prediction of sensitive positions (82.8, 49.8, and 68.5%, respectively) than our evolutionary model and than the two softwares mentioned above, and they also misinterpreted neutral substitutions as being deleterious (9.1, 5.7, and 3.4%, respectively). Therefore, our evolutionary chart of sensitive positions of human TNSALP enables us to validate (or invalidate) at low cost any ALPL mutation, which would be suspected to be responsible for hypophosphatasia. We showed that five in silico models are not as efficient predictive tools as our predictive chart, which allowed us to validate more than 99% of missense mutations in ALPL currently known to result in hypophosphatasia, and did not provide false predictions in contrast to in silico predictive models. We also show that our approach is more efficient in distinguishing precisely causal mutations and polymorphisms, and in avoiding any false prediction (Table 4). Having access to an efficient predictive tool will become more and more necessitated with the dramatic increase of standard and high-throughput sequencing data.

Acknowledgments—We express our sincere thanks to Steve Ursprung and members of the Scientific Committee of Hypophosphatasie Europe for their support and Benjamin Palmer for English editing.

REFERENCES

1. McComb, R. B., Bowers, G. N., and Posen, S. (1979) Alkaline phosphatase, Plenum Press, New York

2. Kam, W., Clauser, E., Kim, Y. S., Kan, Y. W., and Rutter, W. J. (1985) Cloning, sequencing, and chromosomal localization of human term placental alkaline phosphatase cDNA. Proc. Natl. Acad. Sci. U.S.A. 82, 8715–8719

3. Millán, J. L., and Manes, T. (1988) Seminoma-derived Nagao isozyme is encoded by a germ-cell alkaline phosphatase gene. Proc. Natl. Acad. Sci. U.S.A. 85, 3024–3028

4. Berger, J., Garattini, E., Hua, J. C., and Udenfriend, S. (1987) Cloning and sequencing of human intestinal alkaline phosphatase cDNA. Proc. Natl. Acad. Sci. U.S.A. 84, 695–698

5. Yang, Y., Wandler, A. M., Postlethwait, J. H., and Guillemín, K. (2012) Dynamic evolution of the LPS-detoxifying enzyme intestinal alkaline phosphatase in zebrafish and other vertebrates. Front. Immunol. 3, 314

6. Weiss, M. J., Cole, D. E., Ray, K., Whyte, M. P., Lafferty, M. A., Mulivor, R. A., and Harris, H. (1988) A missense mutation in the human liver/bone/kidney alkaline phosphatase gene causes a lethal form of hypophosphatasia. Proc. Natl. Acad. Sci. U.S.A. 85, 7666–7669

7. Henthorn, P. S., Raducha, M., Fedde, K. N., Lafferty, M. A., and Whyte, M. P. (1992) Different missense mutations at the tissue-nonspecific alkaline phosphatase gene locus in autosomal recessively inherited forms of mild and severe hypophosphatasia. Proc. Natl. Acad. Sci. U.S.A. 89, 9924–9928

8. Henthorn, P. S., and Whyte, M. P. (1992) Missense mutations of the tissue-nonspecific alkaline phosphatase gene in hypophosphatasia. Clin. Chem. 38, 2501–2505

9. Greenberg, C. R., Taylor, C. L., Haworth, J. C., Seargeant, L. E., Philips, S., Triggs-Raine, B., and Chodirker, B. N. (1993) A homooligoe Gly317→Asp mutation in ALPL causes the perinatal (lethal) form of hypophosphatasia in Canadian mennonites. Genomics 17, 215–217

10. Mornet, E., Taillandier, A., Peyramare, S., Kaper, F., Muller, F., Brenner, R., Bussière, P., Freisinger, P., Godard, J., Le Merrrer, M., Oury, J. F., Plau- chu, H., Puddu, R., Rival, J. M., Superti-Furga, A., Touraine, R. L., Serre, J. L., and Simon-Bouy, B. (1998) Identification of fifteen novel mutations in the tissue-nonspecific alkaline phosphatase (TNSALP) gene in European patients with severe hypophosphatasia. Eur. J. Hum. Genet. 6, 308–314

11. Mornet, E. (2000) Hypophosphatasia: the mutations in the tissue-nonspecific alkaline phosphatase gene. Hum. Mutat. 15, 309–315

12. Matsuura, S., Kishi, F., and Kaji, T. (1990) Characterization of a 5′-flank-
ing region of the human liver/bone/kidney alkaline phosphatase gene: two kinds of mRNA from a single gene. Biochem. Biophys. Res. Commun. 168, 993–1000.

13. Terao, M., Studer, M., Gianni, M., and Garattini, E. (1990) Isolation and characterization of the mouse liver/bone/kidney-type alkaline phosphatase gene. Biochem. J. 268, 641–648.

14. Studer, M., Terao, M., Gianni, M., and Garattini, E. (1991) Characterization of a second promoter for the mouse liver/bone/kidney-type alkaline phosphatase gene: cell and tissue specific expression. Biochem. Biophys. Res. Commun. 179, 1352–1360.

15. Fedde, K. N., and Whyte, M. P. (1990) Alkaline phosphatase (tissue-nonspecific isoenzyme) is a phosphoethanolamine and pyridoxal 5'-phosphate ectophosphatase: normal and hypophosphatasia fibroblast study. Am. J. Hum. Genet. 47, 767–775.

16. Hoylaerts, M. F., Ding, L., Narisawa, S., Van Kerckhoven, S., and Millán, J. L. (2006) Mammalian alkaline phosphatase catalysis requires active site structure stabilization via the N-terminal amino acid microenvironment. Biochemistry 45, 9756–9766.

17. Le Du, M. H., Stigbrand, T., Taussig, M. J., Menez, A., and Stura, E. A. (2001) Crystal structure of alkaline phosphatase from human placenta at 1.8-Å resolution: implication for a substrate specificity. J. Biol. Chem. 276, 9158–9165.

18. Vittur, F., Stagni, N., Moro, L., and de Bernard, B. (1984) Alkaline phosphatase binds to collagen: a hypothesis on the mechanism of extracellular mineralization in epiphyseal cartilage. Experience 40, 836–837.

19. Wu, L. N., Genge, B. R., Lloyd, G. C., and Wuthier, R. E. (1991) Collagen-binding proteins in collagenase-released matrix vesicles from cartilage. Interaction between matrix vesicle proteins and different types of collagen. J. Biol. Chem. 266, 1195–1203.

20. Bossi, M., Hoylaerts, M. F., and Millán, J. L. (1993) Modifications in a flexible surface loop modulate the isozyme-specific properties of mammalian alkaline phosphatases. J. Biol. Chem. 268, 25409–25416.

21. Weiss, M. J., Henthorn, P. S., Lafferty, M. A., Slaughter, C., Raduca, M., and Harris, H. (1986) Isolation and characterization of a cDNA encoding a human liver/bone/kidney-type alkaline phosphatase. Proc. Natl. Acad. Sci. U.S.A. 83, 7182–7186.

22. Noge, O., Koyama, I., Goseki, M., Roux, B., and Komoda, T. (1997) Human tissue non-specific alkaline phosphatases: sugar-moieties-induced enzymic and antigenic modulations and genetic aspects. Biochem. J. 321, 297–303.

23. Mornet, E. (2008) Hypophosphatasia. Best Pract. Res. Clin. Rheumatol. 22, 113–127.

24. Subramanian, S., and Kumar, S. (2006) Evolutionary anatomy of positions and types of disease-associated and neutral amino acid mutations in the human genome. BMC Genomics 7, 306.

25. Kumar, S., Dudley, J. T., Filipski, A., and Liu, L. (2011) Phylogenetics: an evolutionary telescope to explore and diagnose the universe of disease. Science 334, 521–524.

26. Le Du, M. H., and Weis, K. M. (2003) Mineralized tissue and vertebrate terrestrial revolution and KPg extinction on mammal diversification. Springer, M. S., and Murphy, W. J. (2011) Impacts of the cretaceous terrestrial revolution and KPg extinction on mammalian diversification. Science 334, 521–524.

27. Pond, S. L., Frost, S. D., and Muse, S. V. (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.

28. Pond, S. L., and Frost, S. D. (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol. Biol. Evol. 22, 478–485.

29. Pond, S. L., and Frost, S. D. (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.

30. Betts, M. J., and Russell, R. B. (2003) in Bioinformatics for Geneticists (Barnes, M. R., and Gray, I. C., eds) pp. 289–316, John Wiley & Sons, Ltd., New York.

31. Froussios, K., Sliopoulos, C. S., Schilt, T., and Simpson, M. A. (2013) Predicting the functional consequences of non-synonymous DNA sequence variants: evaluation of bioinformatics tools and development of a consensus strategy. Genomics 102, 223–228.

32. Le Du, M.-H., and Millán, J. L. (2002) Structural evidence of functional divergence in human alkaline phosphatases. J. Biol. Chem. 277, 49808–49814.

33. Stec, B., Holtz, K. M., and Kantrowitz, E. R. (2000) A revised mechanism for the alkaline phosphatase reaction involving three metal ions. J. Mol. Biol. 299, 1303–1311.

34. Kozenkov, A., Manes, T., Hoylaerts, M. F., and Millán, J. L. (2002) Function assignment to conserved residues in mammalian alkaline phosphatases. J. Biol. Chem. 277, 22992–22999.

35. Mornet, E., Stura, E., Lia-Baldini, A. S., Stigbrand, T., Ménez, A., and Le Du, M. H. (2001) Structural evidence for a functional role of human tissue nonspecific alkaline phosphatase in bone mineralization. J. Biol. Chem. 276, 31171–31178.

36. Llanas, P., Stura, E. A., Ménez, A., Kiss, Z., Stigbrand, T., Millán, J. L., and Le Du, M. H. (2005) Structural studies of human placental alkaline phosphatase in complex with functional ligands. J. Mol. Biol. 350, 441–451.

37. Kim, E. E., and Wyckooff, H. W. (1991) Reaction mechanism of alkaline phosphatase based on crystal structures: two-metal ion catalysis. J. Mol. Biol. 218, 449–464.

38. Millán, J. L. (2007) Mammalian Alkaline Phosphatases: From Biology to Applications in Medicine and Biotechnology, John Wiley & Sons, New York.

39. Goseki-Sone, M., Orimo, H., Limura, T., Takagi, Y., Watanabe, H., Taketa, K., Sato, S., Mayanagi, H., Shimada, T., and Oida, S. (1998) Hypophosphatasia: identification of five novel missense mutations (G507A, G705A, A748G, T1155C, G1320A) in the tissue-nonspecific alkaline phosphatase gene among Japanese patients. Hum. Mutat. 1, S263–S267.