

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/236140086>

Molecular Phylogenetic Analyses of Albuminoids Reveal the Molecular Evolution of Allosteric Properties

ARTICLE *in* INTERNATIONAL UNION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY LIFE · JUNE 2013

Impact Factor: 3.14 · DOI: 10.1002/iub.1164 · Source: PubMed

CITATIONS

2

READS

43

6 AUTHORS, INCLUDING:



Loris Leboffe

Università Degli Studi Roma Tre

26 PUBLICATIONS 121 CITATIONS

SEE PROFILE



Tiziana Alberio

Università degli Studi dell'Insubria

29 PUBLICATIONS 250 CITATIONS

SEE PROFILE



Gabriella Fanali

Università degli Studi dell'Insubria

43 PUBLICATIONS 1,156 CITATIONS

SEE PROFILE



Mauro Fasano

Università degli Studi dell'Insubria

169 PUBLICATIONS 4,798 CITATIONS

SEE PROFILE

Research Communication

Molecular Phylogenetic Analyses of Albuminoids Reveal the Molecular Evolution of Allosteric Properties

Paolo Ascenzi^{1,2†}
Alessandra di Masi^{2,3†}
Loris Leboffe^{1,2}
Tiziana Alberio⁴
Gabriella Fanali⁴
Mauro Fasano^{4*}

¹National Institute for Biostructures and Biosystems, Roma, Italy

²Interdepartmental Laboratory for Electron Microscopy, University Roma Tre, Roma, Italy

³Department of Sciences, University Roma Tre, Roma, Italy

⁴Department of Theoretical and Applied Sciences-Biomedical Research Division, University of Insubria, Busto Arsizio (VA), Italy

Abstract

Serum albumin, α -fetoprotein, afamin (also named α -albumin and vitamin E binding protein), and vitamin D binding protein are members of the albuminoid superfamily. Albuminoids are plasma proteins characterized by a marked ability for ligand binding and transport. Here, a focused phylogenetic analysis of sequence evolution by maximum likelihood of fatty acid binding sites FA1–FA7 of mammalian albuminoids reveals that the FA1, FA2, and FA3+FA4 sites in serum albumins have evolved from the most recent common ancestor through an intermediate that has originated the α -fetoprotein and afamin clades. The same topology has been

observed for the whole protein sequences, for the sequences of all the fatty acid binding sites (FA1–FA7) taken together, and for the allosteric core corresponding to residues 1–303 of human serum albumin. The quantitative divergence analysis indicates that the ligand binding cleft corresponding to the FA2 site could be the main determinant of allosteric properties of serum albumins only. In fact, this binding cleft is structurally not effective in vitamin D binding proteins, whereas key residues that serve to allocate the allosteric effectors are not present in afamins and α -fetoproteins. © 2013 IUBMB Life, 00(00):000–000, 2013.

Keywords: afamin; allostery; α -fetoprotein; molecular evolution; serum albumin; vitamin D binding protein

Introduction

Serum albumin (ALB), α -fetoprotein (AFP), afamin (also named α -albumin and vitamin E binding protein; AFM), and vitamin D

binding protein (DBP) are members of the albuminoid superfamily. These proteins display a conserved three-domain organization although with a different architecture and a marked ability for ligand binding and transport. In particular, ALB is made up of three domains (I, II, and III), each domain being constituted by two subdomains (A and B) (1–7). The multidomain organization of ALB is at the root of its capability to bind at multiple sites not only endogenous and exogenous low molecular weight compounds but also peptides and proteins. Notably, ALB is able to bind fatty acids (FAs), its primary physiological ligands, at seven main sites, labeled fatty acid binding site 1–7 (FA1–FA7) (Fig. 1) (6–9,11).

The N-terminal half of ALB, comprising domains I and IIA, represents the allosteric core of the protein, with the FA1, FA2, and FA7 sites being functionally linked (6,12,13). Ligand binding to the FA2 site, identified as the allosteric modulatory cleft, enhances the ligand affinity for the FA1 site and disfavors ligand binding to the FA7 cleft. Moreover, ligand binding to the FA7 site decreases ligand affinity for both FA1 and FA2

Abbreviations: AFM, afamin; AFP, α -fetoprotein; ALB, serum albumin; DBP, vitamin D binding protein; FA, fatty acid

© 2013 International Union of Biochemistry and Molecular Biology, Inc.
Volume 000, Number 000, Month/Month 2013, Pages 000–000

*Address for correspondence to: M. Fasano, Department of Theoretical and Applied Sciences - Biomedical Research Division, University of Insubria, Via Alberto da Giussano 12, I-21052 Busto Arsizio (VA), Italy. Tel: +39 0331 339450. Fax: +39 0331 339459. E-mail: mauro.fasano@uninsubria.it.

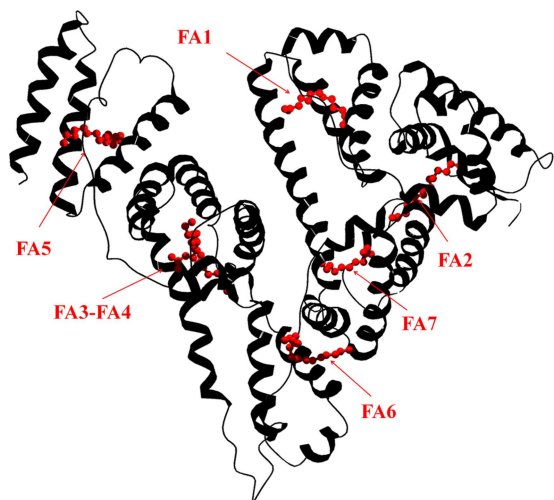
[†]These authors contributed equally to this article.

Received 5 December 2012; accepted 21 February 2013

Additional Supporting Information may be found in the online version of this article

DOI: 10.1002/iub.1164

Published online 00 Month 2013 in Wiley Online Library
(wileyonlinelibrary.com)


FIG 1

Three-dimensional structure of human ALB. The FA binding sites are indicated by myristate ions (in red) rendered in ball and sticks. Atomic coordinates were taken from the PDB ID: 1E7E (9). Structural models were drawn with Swiss-PDBViewer (10). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

clefts. Furthermore, ligand binding to the FA7 site impairs ligand binding to the FA1 cleft and *vice versa* (6,12–17). The C-terminal half of ALB comprises the FA3+FA4, FA5, and FA6 sites that play a relevant role in the binding of FAs and drugs. However, these sites have not been demonstrated to contribute to allosteric properties of ALB, so far (12,13,18). FA binding to the medium-affinity sites FA2 and FA3 drives the conformational transition(s) of ALB. The FA-loaded conformational state of ALB corresponds to the B form of the protein, whereas the FA-free conformational state coincides with its N form (6).

Despite the wealth of information available on ligand binding properties of albuminoids, only information on the molecular evolution of the FA1 site (also named as the heme binding cleft) has been reported. By taking advantage of multiple sequence alignment, three-dimensional structure modeling, and ligand docking simulation, the FA1 site of paralogous human albuminoids AFP and ALB has been hypothesized to have evolved from the ancestor DBP through AFM (19). It appears therefore of interest to examine the conservation of residues in selected sequence stretches corresponding to the FA1–FA7 sites in mammalian albuminoids to get insights on their evolution in terms of ligand binding properties and allosteric modulation (6,9).

Methods

Sequences of mammalian orthologs of AFM, AFP, DBP, and ALB were downloaded from the Universal Protein Resource (uniprot, <http://www.uniprot.org>) (20). Only sequences of proteins with evidence at the mRNA or protein level were

considered. The mature protein sequences were obtained according to uniprot annotations. Progressive multiple alignment was performed using clustalw2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) with BLOSUM (BLOCKS of Amino Acid SUBstitution Matrix) matrices used to score both pairwise and multiple alignments. The putative FA binding sites of AFM, AFP, and DBP were identified in the multiple alignment on the basis of sequence stretches delimiting the FA1–FA7 sites in the human ALB structure (PDB ID:109X), as previously reported (18). The “all sites” multiple alignment was obtained by stitching the FA1–FA7 site sequences together after removing overlapping regions. Similarly, the “excluding sites” multiple alignment was obtained by removing sequence stretches corresponding to the FA binding sites. The multiple alignment was tested with the ProtTest 2.4 server (http://darwin.uvigo.es/software/prottest2_server.html) (21) to select the most appropriate model of protein evolution.

The evolutionary history was inferred by using the Maximum Likelihood method with the MEGA5 software (22). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model (23) and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (five categories, +G). The rate variation model allowed for some sites to be evolutionarily invariable (+I). All positions containing gaps and missing data were eliminated. Trees were drawn to scale, with branch lengths measured in the number of substitutions per site.

Values of divergence from the most recent common ancestor were calculated by submitting final trees to the DistParser server (<http://indra.mullins.microbiol.washington.edu/cgi-bin/DistParser/distcalculate.cgi>) and shown as box-and-whiskers plots using Prism 5 (GraphPad Software, La Jolla, CA). Divergence values were analyzed by the Kruskal-Wallis one-way nonparametric analysis of variance followed by the post hoc Dunn’s multiple comparison test (24). Residue conservation in each position of the multiple alignment was calculated using ConSurf (<http://consurf.tau.ac.il/>) (25). Structural models were drawn with the Swiss-PDBViewer (10) starting from the PDB entries 1E7E and 1E7G for human ALB (8) and 1KW2 for human DBP (26).

Results and Discussion

Analysis of Sequence Alignments for FA1–FA7 Sites

Protein sequences of 16 ALB, three AFM, seven AFP, and five DBP from mammals underwent progressive multiple alignment (Supporting Information Fig. 1). The more appropriate protein evolution model is the JTT+G+I, as identified by the analysis of the multialigned sequences with ProtTest.

Quantitative cladograms were obtained by maximum likelihood for all the sequence regions, *i.e.*, the whole protein, “all sites”, “excluding sites”, the FA1–FA7 binding sites, and subdomains I+IIA (Fig. 2 and Supporting Information Fig. 2). Most

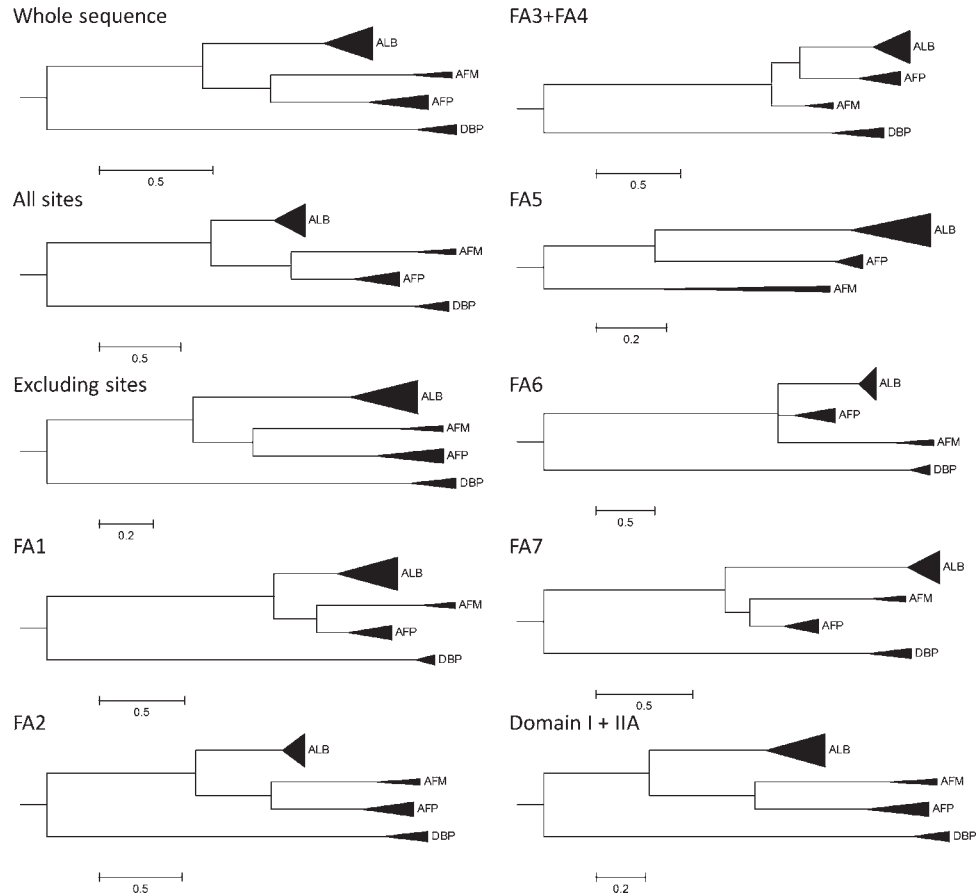


FIG 2

Cladistic trees of ALB, AFP, AFM, and DBP obtained from the whole protein sequences, from sequence stretches corresponding to the ALB FA binding sites, from the sum of all the FA site sequence stretches stitched together ("all sites"), from the whole sequence excluding those of the FA binding sites ("excluding sites"), and from sequences corresponding to residues 1–303 in human ALB (subdomains I + IIA). The definition of the FA site boundaries has been previously reported (18). Cladograms were obtained by maximum likelihood using the JTT+G+I protein evolution model as described in the text. Complete cladograms are reported in Supporting Information Fig. 2.

trees indicate that DBP sequences have evolved from the most recent common ancestor separately from the other members of the albuminoid family. Paralogous sequences appear in different clades, whereas all orthologous sequences are present in the same subtree. This finding strongly suggests that the gene duplication events that gave rise to the four albuminoids occurred before the radiation of mammals. By allowing for some sites to be evolutionarily invariant (2.4% sites in the whole sequence) and by assuming that evolutionary rates may be different among sites, ALB orthologs appear to have evolved separately from AFM and AFP members, that differentiated in a further duplication event. This topology does not agree with the previous hypothesis that AFM should have evolved from an ancestor that subsequently originated the precursors of AFP and ALB (19). It should be noticed, however, that the previously reported phenogram was obtained for the partial 112–199 sequence of human albuminoids, corresponding to the four-helix bundle surrounding the heme in the FA1 site, and should be considered only in terms of a hierarchical clustering

of the conservation of a consensus sequence able to bind the heme (19). The more accurate analysis reported here, including three or more mammalian sequences per protein and using a maximum likelihood approach, is in keeping with the evolutionary analysis recently described in the context of the zebrafish liver differentiation (27).

Worthy of note, the same behavior was not observed when sequence stretches surrounding the FA3+FA4 cleft were considered. In this region, AFP sequences were related more closely to ALB than to AFM sequences. A tentative explanation may be found in the specificity of AFM to bind vitamin E with high affinity (3,4). The moderate affinity of vitamin E for the FA3 site of human ALB (28) might suggest that AFM binds vitamin E in the homologous region. On the other hand, the FA3+FA4 cleft of ALB and AFP might have evolved independently to bind FAs with high affinity, although maintaining a moderate affinity for vitamin E. As far as the FA5 and FA6 sites are concerned, the maximum likelihood phylogenetic analysis did not allow to infer a clear topology to be compared

with the other sequence regions. Moreover, the FA5 site is not defined in DBP, therefore DBP sequences were not considered in the phylogenetic analysis. The FA6 site is characterized by low ligand affinity and specificity (6), leading to high divergence from the most recent common ancestor in all four clades.

To quantitatively evaluate cladograms shown in Fig. 2 and in Supporting Information Fig. 2, divergence values were calculated for each sequence in terms of average substitutions/site according to the JTT+G+I model and their distribution was displayed in terms of box-and-whiskers plots (Fig. 3). The analysis of whole sequences showed a significantly lower divergence of ALB members with respect to the other members. This finding is even more evident when the “all sites” sequences were considered. Accordingly, divergence values were not significantly different when these sequences were excluded from the multiple alignment (“excluding sites”). Each FA binding site displayed similar divergence of ALB with respect to other clades from the most recent common ancestor, with the only exceptions of the FA2, FA3+FA4, and FA7 sites. The significantly lower divergence of the FA3+FA4 cleft in AFM might be ascribed to vitamin E binding specificity, as suggested above. On the contrary, the lower divergence of the FA7 site in AFP does not have, to the best of our knowledge, an unambiguous explanation.

The case of the FA2 site closely resembled the distribution observed for the whole sequence. To further investigate this finding, the sequences of the region identified as the allosteric core in ALB were analyzed, being restricted to domain I and subdomain IIA to focus on the FA1, FA2, and FA7 sites. Subdomains I+IIA behaved in the same way as observed for the whole protein, the FA2 site, and the “all sites” sequences. As a whole, FA2 should be responsible for the reduced divergence of ALB sequences with respect to those of AFM, AFP, and DBP, thus suggesting that the peculiar allosteric regulation properties associated to the FA2 site should be a prerogative of ALB.

Analysis of Sequence and Structural Variability of the FA2 site

All the FA binding sites show appreciable sequence variability in albuminoids. A complete description of the residue conservation in each position is reported in Supporting Information Table 1. Nevertheless, we focused on residues that play a functional role in the FA2 site to understand the bases of the lower divergence of ALB described above (Fig. 3).

The FA2 site is one of the most internal FA binding clefts of ALB, being located between subdomains IA and IIA. Within this site, we examined the variability of residues Tyr150, Arg257, and Ser287 that stabilize the FA carboxylate group by hydrogen bond interactions (9,29). In the FA2 site, the human ALB residue Tyr150 is conserved in ALB and AFP only, being Phe in AFM and Pro in DBP. Conversely, Arg257 is conserved only in ALB. Also Ser287 is not conserved; in fact, Ser is replaced by Ala in mouse and rabbit ALB, whereas in the other albuminoids this position is occupied by Glu, Phe, Gly, and Ala. This suggests that FA2 is a medium-affinity site in ALB only. Reasonably, the replacement of Ser by Ala in mouse

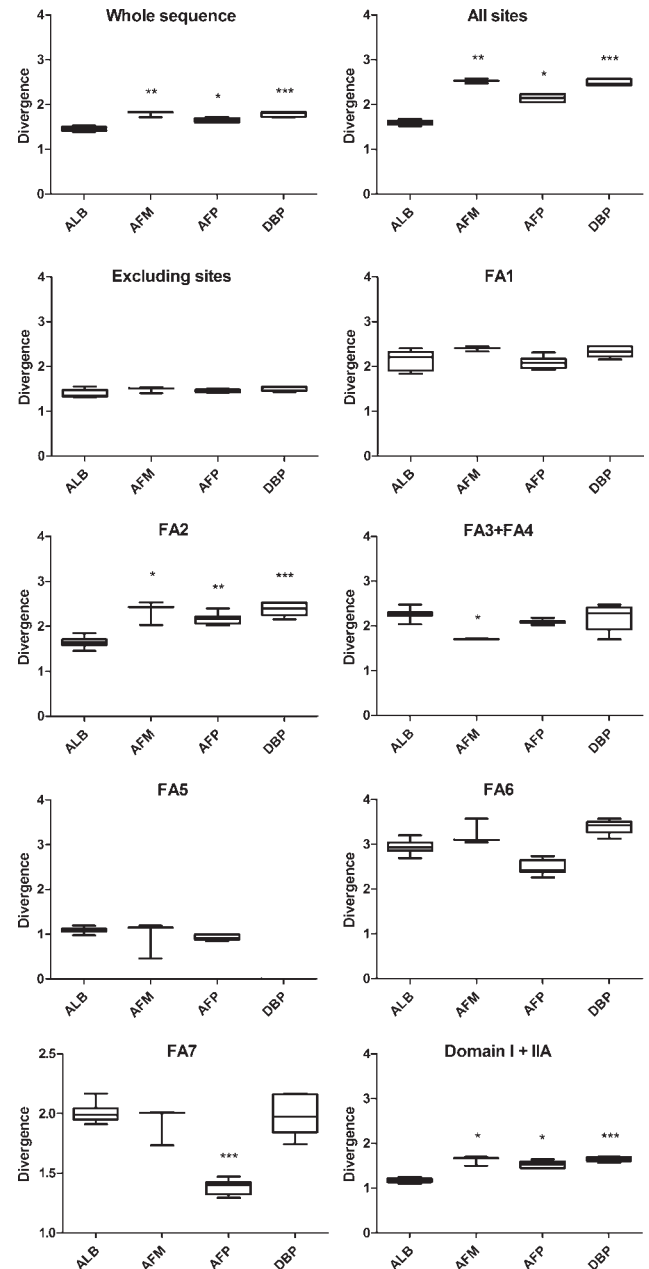


FIG 3

Values of divergence from the most recent common ancestor (measured as average substitutions/site) for the clades in Fig. 2. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$ (Kruskal-Wallis test followed by Dunn's post hoc test).

and rabbit ALB should not affect dramatically the FA2 binding affinity, Tyr and Arg being conserved (Fig. 4). Accordingly, the allosteric effect that follows FA2 ligand binding could be an exclusive peculiarity of ALB.

Only FA sites of ALB located in domains I and II have been reported to be allosterically coupled, so far. Indeed, the recombinant Asp1-Glu382 truncated ALB, encompassing only domains I and II, displays spectroscopic and reactivity properties

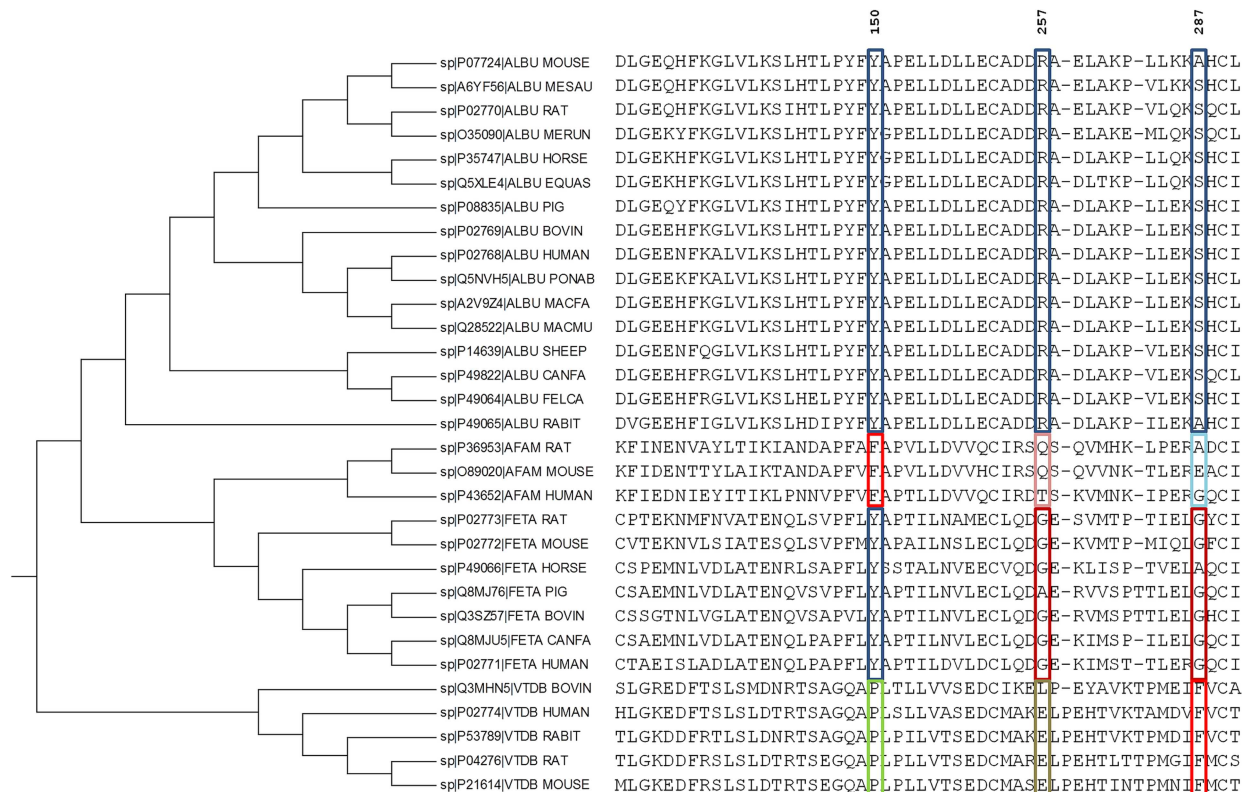


FIG 4

Evolution of the FA2 site. Clustered sequences in close contact with the FA2 ligand are shown, with residues corresponding to human ALB Tyr150, Arg257, and Ser287 positions being highlighted. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

superimposable to those of the full-length protein and represents a valuable model to investigate the allosteric linkage between the FA1, FA2, and FA7 sites (30,31). Figure 5 reports the allosteric core of ALB superimposed with the homologous region of human DBP (PDB ID: 1KW2) (26). Although domain I could be superimposed, subdomain IIA was rotated around the interdomain helix by about 90°. As a consequence, the three residues that stabilize the polar head of the FA anion bound in the FA2 site contribute to build the cavity that hosts the ligand only in ALB, whereas in DBP this cleft is no longer defined.

As far as AFP and AFM are concerned, no structural data are available at present, therefore we cannot speculate about the conformation of the FA2 site in these albuminoids. It appears reasonable, however, that the absence of the three key residues Tyr, Arg, and Ser should impair the affinity of FA anions for this site, so that the physiological FA concentration could not be able to switch the allosteric transition.

Conclusion

The analysis of maximum likelihood cladograms of albuminoids indicates a reduced divergence of ALB with respect to other members that may be ascribed to the conservation of the

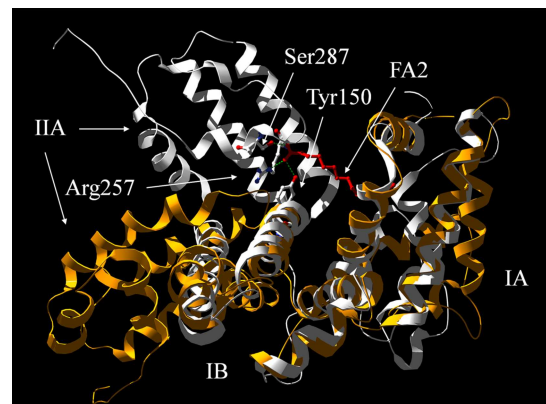


FIG 5

The N-terminal allosteric region of human ALB, corresponding to the subdomains I + IIA (residues 1–303) (in white ribbons), compared with the corresponding region in human DBP (in yellow ribbons). Residues Tyr150, Arg257, and Ser287 are drawn in ball-and-sticks. Dashed green lines indicate hydrogen bonds with the polar head of the myristate anion (in red) in the FA2 site. Atomic coordinates for human ALB and DBP were taken from the PDB ID: 1E7G (9) and 1KW2 (26), respectively. Structural models were drawn with Swiss-PDBViewer (10). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

FA2 site. In keeping with this finding, the three key residues Tyr150, Arg257, and Ser287, endowing the FA2 cavity with medium affinity for FA anions, are conserved in ALB only. Additionally, the region homologous to FA2 in DBP does not form a cavity able to host FAs. As a whole, the allosteric properties linked to the FA2 site should be a prerogative of ALB.

Acknowledgements

The authors thank Prof. Giorgio Bernardi for the helpful discussion.

References

- [1] Sugio, S., Kashima, A., Mochizuki, S., Noda, M., and Kobayashi, K. (1999) Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* 12, 439–446.
- [2] Mizejewski, G. J. (2001) Alpha-fetoprotein structure and function: relevance to isoforms, epitopes, and conformational variants. *Exp. Biol. Med.* 226, 377–408.
- [3] Voegelé, A. F., Jerkovic, L., Wellenzohn, B., Eller, P., Kronenberg, F., et al. (2002) Characterization of the vitamin E-binding properties of human plasma afamin. *Biochemistry* 41, 14532–14538.
- [4] Jerkovic, L., Voegelé, A. F., Chwatal, S., Kronenberg, F., Radcliffe, C. M., et al. (2005) Afamin is a novel human vitamin E-binding glycoprotein. Characterization and in vitro expression. *J. Proteome Res.* 4, 889–899.
- [5] Speckaert, M., Huang, G., Delanghe, J. R., and Taes, Y. E. (2006) Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin. Chim. Acta* 372, 33–42.
- [6] Fanali, G., di Masi, A., Trezza, V., Marino, M., Fasano, M., et al. (2012) Human serum albumin: from bench to bedside. *Mol. Aspects Med.* 33, 209–290.
- [7] Ahmad, E., Rabbani, G., Zaidi, N., Ahmad, B., and Khan, R. H. (2012) Pollutant-induced modulation in conformation and β -lactamase activity of human serum albumin. *PLoS One* 7, e38372.
- [8] Peters, T. Jr. (1996) All about Albumin: Biochemistry, Genetics and Medical Applications. Academic Press, San Diego, CA.
- [9] Bhattacharya, A. A., Grüne, T., and Curry, S. (2000) Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *J. Mol. Biol.* 303, 721–732.
- [10] Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30 (Suppl. 1), S162–S673.
- [11] Fasano, M., Curry, S., Terreno, E., Galliano, M., Fanali, G., et al. (2005) The extraordinary ligand binding properties of human serum albumin. *IUBMB Life* 57, 787–796.
- [12] Ascenzi, P. and Fasano, M. (2009) Serum heme-albumin: an allosteric protein. *IUBMB Life* 61, 1118–1122.
- [13] Ascenzi, P. and Fasano, M. (2010) Allostery in a monomeric protein: the case of human serum albumin. *Biophys. Chem.* 148, 16–22.
- [14] Chuang, V. T. and Otagiri, M. (2006) Stereoselective binding of human serum albumin. *Chirality* 18, 159–166.
- [15] Fanali, G., Fesce, R., Agrati, C., Ascenzi, P., and Fasano, M. (2005) Allosteric modulation of myristate and Mn(III)-heme binding to human serum albumin: optical and NMR spectroscopy characterization. *FEBS J.* 272, 4672–4683.
- [16] Simard, J. R., Zunszain, P. A., Hamilton, J. A., and Curry, S. (2006) Location of high and low affinity fatty acid binding sites on human serum albumin revealed by NMR drug-competition analysis. *J. Mol. Biol.* 361, 336–351.
- [17] Fanali, G., Cao, Y., Ascenzi, P., Trezza, V., Rubino, T., et al. (2011) Binding of δ 9-tetrahydrocannabinol and diazepam to human serum albumin. *IUBMB Life* 63, 446–451.
- [18] Fanali, G., Ascenzi, P., Bernardi, G., and Fasano, M. (2012) Sequence analysis of serum albumins reveals the molecular evolution of ligand recognition properties. *J. Biomol. Struct. Dyn.* 29, 691–701.
- [19] Fasano, M., Fanali, G., Leboffe, L., and Ascenzi, P. (2007) Heme binding to albuminoid proteins is the result of recent evolution. *IUBMB Life* 59, 436–440.
- [20] Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, B., et al. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* 10, 136.
- [21] Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105.
- [22] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- [23] Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- [24] McDonald, J. H. (2009) Handbook of Biological Statistics, 2nd ed. Sparky House Publishing, Baltimore, MD.
- [25] Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38, W529–W533.
- [26] Otterbein, L. R., Cosio, C., Graceffa, P., and Dominguez, R. (2002) Crystal structures of the vitamin D-binding protein and its complex with actin: structural basis of the actin-scavenger system. *Proc. Natl. Acad. Sci. USA* 99, 8003–8008.
- [27] Noël, E. S., Reis, M. D., Arain, Z., and Ober, E. A. (2010) Analysis of the Albumin/ α -Fetoprotein/Afamin/Group specific component gene family in the context of zebrafish liver differentiation. *Gene Expr. Patterns* 10, 237–243.
- [28] Fanali, G., Fasano, M., Ascenzi, P., Zingg, J. M., and Azzi, A. (2013) α -Tocopherol binding to human serum albumin. *Biofactors*; doi: 10.1002/biof.1070.
- [29] van der Vusse, G. J. (2009) Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.* 24, 300–307.
- [30] Fanali, G., Pariani, G., Ascenzi, P., and Fasano, M. (2009) Allosteric and binding properties of Asp1-Glu382 truncated recombinant human serum albumin: an optical and NMR spectroscopic investigation. *FEBS J.* 276, 2241–2250.
- [31] di Masi, A., Gullotta, F., Bolli, A., Fanali, G., Fasano, M., et al. (2011) Ibuprofen binding to secondary sites allosterically modulates the spectroscopic and catalytic properties of human serum heme-albumin. *FEBS J.* 278, 654–662.