Abstract—Aiming the isolation and molecular characterization of the gene for the d polypeptide chain of giant extracellular hemoglobin of Glossoscolex paulistus (HbGp), we present the design of oligonucleotides to DNA hybridization procedures. From a described amino acid sequence, we performed reverse translation, with an algorithm, which searched for regions with maximum density of non-degenerate codons, focusing minimization of the number of putative oligonucleotide probes. Fourteen different partial oligopeptide sequences six, seven or eight residues long resulted in 64, 256 and 512 possible coding sequences, respectively. Based on the similarity with d chain of the hemoglobin from Lumbricus terrestris (HbLt), we ranked the resulting sequences. This narrowed the candidate sequences to 2, 20 and 24 possible probes, favoring shorter sequences. Thus, five sequences targeting two different regions and possible primer pairs for PCR with the HbGp d chain gene were obtained. These oligonucleotides can be employed in hybridization and PCR essays in Glossoscolex paulistus genomic DNA.

Index Terms—Extracellular giant hemoglobin, Glossoscolex paulistus, HbGp, Lumbricus terrestris, reverse translation.

I. INTRODUCTION

The giant extracellular hemoglobins constitute the summit of structural complexity in hemoproteins that carries molecular oxygen [1]-[3]. These supramolecular systems present a very interesting quaternary structure, involving great number of polypeptide subunits. This great number of polypeptide chains favor the hydrophobic isolation of the prosthetic groups, i.e., the heme groups, which constitute the active site of these hemoproteins, being directly responsible to the bond and carry the oxygen molecule. This isolation is a fundamental prerequisite to avoid significant water accessibility to the heme pocket, which is directly associated to the oxidation of the ferrous ion to ferric ion, which is not a functional coordination center to bind the oxygen ligand [4]-[6].

Great number of important works has been published focused on the giant extracellular hemoglobin of Lumbricus terrestris (HbLt), which has been considered relevant reference in this area of research [7]-[9]. We here focus on the giant extracellular hemoglobin of Glossoscolex paulistus (HbGp), which presents significant similarities regarding redox, structural and physico-chemical properties in relation to HbLt. Glossoscolex paulistus is an annelid found in a considerable quantity in the state of São Paulo, Brazil, mainly in the regions of the municipalities of Rio Claro (SP) and Piracicaba (SP) [10].

Giant extracellular hemoglobin from Glossoscolex paulistus (HbGp) is a multi-subunit protein relevant in diverse areas, including biochemistry, biotechnology, medicine and evolution. This giant extracellular hemoglobin presents high resistance to autoxidation, thus being prototype to biosensors and blood substitutes. This hemoglobin constitutes a quaternary arrangement with 180 polypeptide subunits, being 144 polypeptide chains with heme group (globin chains) and 36 polypeptide chains without heme group, which are called “Linker chains”, similarly to the HbLt [11]. These “Linker chains” present only structural function, since they cannot carry oxygen molecules because does not have the ferrous ion-porphyrin system to bind this ligand. The “Linker chains” are disposed more predominantly in the central part of the hexagonal quaternary arrangement, being associated, at least partially, to the cohesion of the supramolecular structure.

The spatial arrangement of HbGp comprises two mutually superposed hexagons, sometimes referred to as a “bracelet”, due its complex stereochemical disposition, generating hexagonal bilayer that involves a total supramolecular mass of 3.6 MDa [11]. Each vertex of each hexagon has three abcd tetradiemers (constituting a dodecamer each). Considering that the two hexagonal arrangements presents a total value of 12 dodecamers, there is a total of 144 globin chains, since the polypeptide chains a, b, c, and d are different types of globin polypeptides [12].

In this context, it is important to notice that the chains a, b and c constitute a trimer (trimer abc), through disulfide bonds and the monomer d is maintained in the tetramer with only weak interaction, i.e., without covalent bonds, like, for example, disulfide bonds. In this way, the study of the monomer d has been considered an important resource of information on the tertiary structure, since the monomer d is the unique polypeptide subunit that can be easily isolated from the whole quaternary structure [13].

Érico M. Polo, Rosiane P. Santos, Leonardo M. Moreira, and Gabriel M. Yazbeck

Bioinformatical Development of Oligonucleotides for the d Chain Gene of the Giant Extracellular Hemoglobin of Glossoscolex Paulistus

Manuscript received August 15, 2013; revised November 13, 2013. This work was supported in part by Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) and Companhia Eletrica de Minas Gerais (CEMIG). Érico Macedo Polo is with the Genetics, Conservation and Evolutionary Biology Program (PPG-GCBEV), Instituto Nacional de Pesquisas da Amazônia (INPA), located at Av. André Araújo, 2936 – Aleixo - Manaus/AM, CEP 69011-970 (e-mail: erico.polo@gmail.com).

Rosiane de Paula Santos, Leonardo Marmo Moreira, and Gabriel Menezes Yazbeck are with the Laboratório de Recursos Genéticos, DEZOO, Universidade Federal de São João Del-Rei, São João Del Rei, MG, CEP 36307-352, Brazil (e-mail: rosianneps2007@yahoo.com.br, leonardomarmo@gmail.com, dna@ufmg.br).

DOI: 10.7763/IJBBB.2014.V4.312

This kind of hemoglobin present high resistance to the autoxidation, which is deeply associated to its complex spatial arrangement [14]. Indeed, these hemoglobins are disposed directly in the blood, without any erythrocytes, differently of the mammalian hemoglobins, and even in these drastic chemical environment conditions present great redox stability. This can be easily observed through of its redox and structural stabilities even when submitted to storage procedures without control of the oxygen presence, i.e., without conditions of inert atmosphere. [15]. Furthermore, these hemoglobins are considered hemoprotein systems with great cooperativity, and in some cases with “supercooperativity” [16], [17].

The affinity of the ferrous ion for molecules of great biological and environmental relevance makes this giant extracellular hemoglobin an interesting model as a “biosensor”. Indeed, compounds as nitric oxide (NO), carbon monoxide (CO), cyanide anion (CN-), among others, are examples of ligands that present high chemical stability when coordinated to the ferrous ion [4]. This happens because these compounds are highly π-acid ligands (π-acceptor ligands), which present significant affinity to the ferrous ion (Fe²⁺), which, in turn, present d⁶ low-spin electronic configuration. Therefore, in the d orbitals with pi symmetry, this cationic center present high electronic density, being more effectively stabilized when bounded with π-receptor ligands. In this way, this hemoglobin could be the main component of new materials elaborated in order to detect important molecules in several biological and non-biological environments.

Motivated by the lack of availability of the gene sequence for the HbGp d chain, we herein approached a reverse translation procedure of a 142 residues long protein sequence in order to generate a set of putative oligonucleotide probes and PCR primers, suitable for molecular essays with genomic DNA from G. paulistus.

II. MATERIAL AND METHODS

We developed an original algorithm (written in Perl language, available upon request from author EMP) that used the HbGp d chain aminoacid sequence as input. It attributes to each position a value which is given by the number of different alternative codons for that particular aminoacid, using the standard universal genetic code [18]. Then, given a fixed arbitrary size class of aminoacid residues (six, seven or eight), the program would return the regions with the maximum density of non-degenerate or twofold degenerate codons. Through the use of the original algorithm written here, we obtained all possible 18 to 24 bp DNA sequences. These were recorded in FASTA format and used to perform a nucleotide BLAST search [19] in GenBank. Then, the higher scoring sequences were compared and rescored in relation to each other, regarding relative codon usage frequency, estimated from the HbLt d chain data. Sequences with the five highest scores from each size class were selected. This procedure favored the maintenance of shorter sequences, every time two sequences overlapped or embedded one another. Oligonucleotide probes and PCR primers were designed from the resulting sequences. Oligonucleotide properties were theoretically derived with the aid of the Oligonucleotide Properties Calculator [20].

III. RESULTS

A total of 14 different oligopeptide stretches six, seven or eight residues long (three, seven and four, respectively) were selected, based on their degenerency density, resulting in the back translation of 64, 256 and 512 possible underlying coding nucleotide sequences, respectively, binding to three different regions of the analyzed protein (Fig. 1).
IV. DISCUSSION

The oligonucleotide probe pool herein proposed is the first set of putative candidates capable of being used in molecular essays involving the use of *G. paulistus* genomic DNA for the screening of the *d* chain gene *in vitro*. The set was strictly developed using an *in silico* approach and the steps used here to narrow down the size of the universe of possibilities could be generalized to be applied to other molecular systems. Even though none of the proposed possibilities might be the real one, this procedure allows for an “educated guess” about the putative true underlying coding sequence in *G. paulistus*. The accuracy of this technique could be accessed in the future by means of computer simulations and application to data sets of known protein to nucleic acid correspondence.

Also, the empirical use of the proposed probes could be used to validate the results found here. The application of the designed probes, under a set of varying hybridization stringency conditions (i.e. different salt concentrations and annealing temperatures) will allow for the further refinement and scrutinizing among the pool. The two 18 bases long probes target anchor in the last selected region and the three larger 21 bases probes will target the middle region. This allowed for the proposal of PCR primers to be designed from the oligonucleotide pool. An *in silico* PCR simulation with one herein proposed primers pair and the the HbLt, allowing for 50% of degeneracy predicted the formation PCR product, but caution must be taken since the *L. terrestris* sequence used here for comparison [22] was derived from a cDNA originated from reverse transcription from the *d* chain mRNA, which lacks intron sequences. Genomic essays using this primers will most likely encompass a larger amplicon, due to the presence of possible introns, usually two, interrupting three exon regions in *L. terrestris* globins [22].

The isolation of this particular globin chain from *G. paulistus* is of great relevance for biotechnology and can provide an *in vivo* recombinant system for bacterial expression of the *d* chain peptide in large scales, thus facilitating the access to this material that can be further explored to its potential application as a biosensor.

Other very auspicious applications of giant extracellular hemoglobin is focused on the elaboration of “blood substitute”, which are also denominated “artificial blood” [23] and the relation between the presence of hemoproteins and the generation and/or control of free radical generation. In fact, the problem of blood transfusion constitutes a great academic challenge in all world, since high number of persons dye all years due to the absence of enough quantity of suitable blood to a rapid transfusion. Regarding the studies focused free radical generation, this approach has received great attention of relevant research groups due to the association between the hemoproteins and the free radicals generations, which are responsible for several biological processes, including physiological and pathological chemical mechanisms. Furthermore, it is interesting to mention the studies upon hybrid hemoproteins [24], [25] and reconstitution of hemoproteins [26], [27] has furnished a representative advancement in the understanding of the structure-activity relationship of very complex metalloproteins.

Several relevant research areas associated to the possible applications of giant hemoglobins, such as biomedical engineering, materials engineering, supramolecular chemistry, between others, require a more detailed characterization of these proteins, aiming a consistent understanding upon its structure-function relationship. In this way, advancements in the characterization in terms of molecular biology can be a very interesting alternative to obtain more detailed physico-chemical and biochemical information of this complex supramolecular system.

V. CONCLUSION

The present work constitutes a relevant and original contribution to the understanding of the molecular bases of the monomer *d* of HbGp. This approach, involving an emphasis related to genetics and molecular biology, is an important prerequisite to the isolation and full characterization of its coding gene. This will facilitate the development of expression systems capable of yielding large quantities of this polypeptide subunit and thus helping in the elucidation of several physico-chemical properties of the chain *d*. In this way, aspects involving the evolution of the giant extracellular hemoglobins could be more effectively studied. We believe that this present work can be a relevant contribution to obtain a more consistent analysis of spectroscopic data inherent to giant extracellular hemoglobins, which have been obtained from several research groups. These novel findings would be decisive in the development of various potential biotechnological applications of HbGp. New evaluations regarding monomer *d* of HbGp has been made in our group and soon will be published.

REFERENCES


