Classifying lipoproteins based on their polar profiles

Carlos Polanco1,2, Jorge Alberto Castañón-González3, Thomas Buhse3, Vladimir N. Uversky4,5,6 and Rafael Zonana Anmke7

1Department of Mathematics, Faculty of Sciences, Universidad Nacional Autónoma de México, C.P. 04510 DF, México; 2Departments of Critical Care Medicine and Biomedical Research, Hospital Juárez de México, D.F. 07760 México; 3Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, C.P. 62209 Chamilpa, Cuernavaca, Morelos, México; 4Department of Molecular Medicine and USF Health Byrd Alzheimer’s Research Institute, Morsani College of Medicine, University of South Florida, MDC07 Tampa, FL 33647, USA; 5Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia; 6Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia; 7Faculty of Health Sciences, Universidad Anáhuac, C.P. 52786 Huixquilucan Estado de México, México

The lipoproteins are an important group of cargo proteins known for their unique capability to transport lipids. By applying the Polarity index algorithm, which has a metric that only considers the polar profile of the linear sequences of the lipoprotein group, we obtained an analytical and structural differentiation of all the lipoproteins found in UniProt Database. Also, the functional groups of lipoproteins, and particularly of the set of lipoproteins relevant to atherosclerosis, were analyzed with the same method to reveal their structural preference, and the results of Polarity index analysis were verified by an alternate test, the Cumulative Distribution Function algorithm, applied to the same groups of lipoproteins.

Key words: potential of association; lipoproteins; polar balance; polarity index method; unfolded proteins; folded proteins; partially folded proteins; intrinsically disordered proteins; atherosclerosis

INTRODUCTION

Studies conducted in the 1980s documented a direct correlation between the high blood lipid levels and atherosclerosis (Guyton & Klemp, 1989). They also demonstrated that lowering lipid levels in the blood was associated with a reduction in the cardiovascular disease-related events and atherosclerosis. Parallel studies showed that lipoproteins; i.e., special particles containing both proteins and lipids bound to proteins, represent an important form of lipids transportation in aqueous media (Kostner, 1983; Morrisett et al., 1975; Scanu & Wisdom, 1972).

Lipoproteins differ in the protein to lipids ratio and, in particular, in apolipoproteins and lipids that they contain. For example, the major apolipoproteins include apoE, apoB, apoA-I, apoA-II, apoA-IV, apoC-I, apoC-II, and apoC-III (Mahley et al., 1984). It was also pointed out that being involved in the transport and redistribution of lipids among various cells and tissues, specific apolipoproteins contribute significantly to the regulation of lipoprotein metabolism (Mahley et al., 1984). Based on their density defined by the protein to lipids ratios, lipoproteins are grouped into six classes: Chylomicrons, Very Low Density Lipoproteins (VLDL), Intermediate Density Lipoproteins (IDL), Low Density Lipoproteins (LDL), High Density Lipoproteins (HDL), and lipoproteins relevant to Atherosclerosis (Atherosclerosis) (Table 1) (Garrett & Grisham, 2012; Koba et al., 2003).

In order to deepen the understanding of this important group of cargo-proteins specialized in the transport of lipids, we applied the supervised computational tool called Polarity index method (Polanco et al., 2012), which has already been used in the analysis of other groups of peptides and proteins (Polanco & Samaniego, 2009; Polanco et al., 2012, 2013; 2013a; 2014a; 2014b; 2014c; 2014d; 2014e). Here, this technique was used to characterize Chylomicrons, VLDLs, IDLs, LDLs, HDLs, and Atherosclerosis, i.e. large VLDL, small dense LDL, and small DHDL subclasses (Koba et al., 2003). Polarity index method performed effective identification of the above mentioned groups of lipoproteins based on the records of the 16 possible polar incidents chosen from the four polar groups: polar positively charged (P+), polar negatively charged (P−), polar neutral (N), and non-polar (NP). This approach can also associate some structural features (e.g., the intrinsic disorder of the proteins) with a group of proteins. This work aims at a comprehensive analysis of the cargo-proteins, even though it is known that fragments of these proteins possess different structural profiles, and that structures of apolipoproteins can change due to the binding and release of lipids. The calibration of the polarity index method was done with the entire set of human lipoproteins (full-length proteins and their fragments, Table 1) downloaded from UniProt Database (Magrane, 2011) and split into the mentioned categories. These same proteins were checked for their structural classifications (Table 2): unfolded, partially folded, or folded based on the correlation with the corresponding protein sets from Oldfield et al. (Supplementary material, Oldfield et al., 2005). The results of polarity index method were verified with an alternate test, by applying a Cumulative Distribution Function algorithm over the same groups of lipoproteins (Dunker et al., 2000).

MATERIAL AND METHODS

The latest version of the polarity index method (Polanco et al., 2012) is fully automated.
Intermediate density lipoprotein

Three tests were performed to measure the frequency of each lipoprotein group (Fig. 1). Figure 1 illustrates that such distributions correlate well with the results in Table 3, showing that the locations of the inflection points in the x-axis do not match for the groups. Although the correlation between the lipoprotein groups with structural features (Table 4) is not definitively conclusive, because of the relatively low efficiency of 55% on the scale of 100%, this analysis suggests that all the lipoprotein groups have the folded-partially folded profile. The affinity of the structural features to the lipoprotein groups (Table 5) corroborates the aforementioned finding that these proteins are characterized by the folded-partially folded profile. The Cumulative Distribution Function analysis (see Materials & Methods, Test plan) did not reach a definitive conclusion on the structural assignment of lipoproteins.

**DISCUSSION**

Since the experimental evidence (Kay et al., 1982) correlates various diseases to structural characteristics of lipoproteins, our work brought in an analytical verification for the polar differentiation of the subgroups studied. These polar differences are not observable in the points of maximum and minimum (non-degenerated singularities) (Thom, 1952), but are evidenced by variability of the location of the inflection points (degenerated

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Symbol</th>
<th>Search engine</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>High-density lipoprotein</td>
<td>HDL</td>
<td>(&quot;high density lipoprotein&quot; AND &quot;homo sapiens&quot;) AND hdl AND reviewedyes NOT &quot;lipoprotein binding protein&quot; NOT (disease:&quot;high density lipoprotein&quot;)</td>
</tr>
<tr>
<td>0</td>
<td>Intermediate density lipoprotein</td>
<td>IDL</td>
<td>(&quot;intermediate density lipoprotein&quot; AND &quot;homo sapiens&quot;) AND idl AND reviewedyes NOT &quot;lipoprotein binding protein&quot; NOT (disease:&quot;intermediate density lipoprotein&quot;)</td>
</tr>
<tr>
<td>51</td>
<td>Low density lipoprotein</td>
<td>LDL</td>
<td>(&quot;low density lipoprotein&quot; AND &quot;homo sapiens&quot;) AND ldl AND reviewedyes NOT &quot;lipoprotein binding protein&quot; NOT (disease:&quot;low density lipoprotein&quot;)</td>
</tr>
<tr>
<td>17</td>
<td>Very low density lipoprotein</td>
<td>VLDL</td>
<td>(&quot;very low density lipoprotein&quot; AND &quot;homo sapiens&quot;) AND vldl AND reviewedyes NOT &quot;lipoprotein binding protein&quot; NOT (disease:&quot;very low density lipoprotein&quot;)</td>
</tr>
<tr>
<td>14</td>
<td>Chylomicrons</td>
<td>Chylomicrons</td>
<td>(&quot;chylomicrons&quot; AND &quot;homo sapiens&quot;) AND chylomicrons AND reviewedyes NOT &quot;lipoprotein binding protein&quot; NOT (disease:&quot;chylomicrons density lipoprotein&quot;)</td>
</tr>
<tr>
<td>15</td>
<td>Lipoproteins relevant to atherosclerosis</td>
<td>Atherosclerosis</td>
<td>(lipoproteins atherosclerosis AND organism:&quot;Homo sapiens (Human)&quot; [9606])</td>
</tr>
</tbody>
</table>

**Metrics.** The polarity index is a supervised type method of the Quantitative Structure Activity Relationship (QSAR), which evaluates a single physiochemical property, the Polarity. Its metric requires data training (Tables 1, 2). Each data training set consists of sequences of amino acids which were previously converted to their numerical equivalent according to the rule: \{P+, P-, N, NP\}: P- = \{D, E\}, P+ = \{H, K, R\}, NP = \{A, F, I, L, M, P, V, W\}, and N = \{C, G, N, Q, S, T, Y\} (Timberlake, 1992). Each pair of amino acids in the sequence is counted and registered in an incident matrix; i.e., (row, column) = (amino acid A, amino acid B). These pairs of amino acids are formed when reading the amino acid sequence of each protein from N-terminus to C-terminus (from left to right equivalently), moving one amino acid at a time. The incident matrix (from the training data) is compared with the corresponding matrix of each target sequence. Those sequences that score greater than the default percentage (Table 3), are considered to be candidate proteins.

**UniProt database preparation.** 101 lipoproteins present in Homo sapiens grouped in six classes were downloaded from the UniProt database (Magrane, 2011) on September 30, 2014 (Table 1). We verified that each of the proteins exists in only one of the sets. This restriction removed the group of Intermediate Density Lipoprotein (IDL), so only five types of lipoproteins were studied.

**Supplementary material.** 352 fragments with known structural attributes that were properly annotated in the supplementary material (Table 2) of Oldfield et al. (Oldfield et al., 2005) were considered.

**Test plan.** Three tests were performed to measure the two following aspects: the property of being cargo-protein, and the level of structural disorder.

The polarity index method (Polanco et al., 2012) calibrated with every lipoprotein group (Table 1) in order to: (i) measure the number of hits during the identification of lipoproteins (Table 3), and (ii) plot the relative frequency of each lipoprotein group (Fig. 1).

The polarity index method calibrated with each group of ordered, disordered and partially disordered fragments (Table 2) in order to measure: (i) the correlation between the Lipoproteins groups and groups of various disorders (Table 4), and, vice versa, (ii) the correlation between the disorder parameters and the lipoprotein groups (Table 5).

Points 2 and 3 show that the polarity index method is a bijection; i.e., it finds correlations between Group A and Group B, and between Group B and Group A.

The Cumulative Distribution Function algorithm (Dunker et al., 2000) calibrated with the disordered fragments in order to be applied on the lipoprotein groups.

**RESULTS**

The polarity index method (Polanco et al., 2012) identified, with a high efficiency of 75%, each of the group of Lipoproteins (Table 3). It also provided means to plot the distribution of relative frequencies of each group (Fig. 1). Figure 1 illustrates that such distributions correlated well with the results in Table 3, showing that the locations of the inflection points in the x-axis do not match for the groups. Although the correlation between the lipoprotein groups with structural features (Table 4) is not definitively conclusive, because of the relatively low efficiency of 55% on the scale of 100%, this analysis suggests that all the lipoprotein groups have the folded-partially folded profile. The affinity of the structural features to the lipoprotein groups (Table 5) corroborates the aforementioned finding that these proteins are characterized by the folded-partially folded profile. The Cumulative Distribution Function analysis (see Materials & Methods, Test plan) did not reach a definitive conclusion on the structural assignment of lipoproteins.

**DISCUSSION**

Since the experimental evidence (Kay et al., 1982) correlates various diseases to structural characteristics of lipoproteins, our work brought in an analytical verification for the polar differentiation of the subgroups studied. These polar differences are not observable in the points of maximum and minimum (non-degenerated singularities) (Thom, 1952), but are evidenced by variability of the location of the inflection points (degenerated
sing singularities), where the locations of all of these points are not matching between different curves. The polarity index method was used to measure all these sequence-based differences and, because of this, exhibited a high level of efficiency (75–86%) in the identification of the subgroups studied. We think that the efficiency of this method was due to the comprehensive nature of its metric that considers the 16 possible polar interactions and does not assess the quality of the polarity of the protein with a single number. Such use of 16 measures and not only one provides more comprehensive evaluation of the polarity of the protein.

Although finding the degree of intrinsic disorder (Uversky, 2002) of proteins provides a useful means for protein analysis, we did not find a reliable correlation between the structural assignment of a given protein to ordered, unfolded, or partially ordered structural categories and its classification to a given lipoprotein class. There are results (Knowles et al., 2014; Uversky & Fink, 2004; Uversky, 2009; 2010; 2014) suggesting that the degree and/or proportion of intrinsic disorder of the proteins are associated with increased risk of amyloidosis (Pepys, 2006). The amyloidogeneity of a protein can be typically assigned to its some specific fragments (Pawlicki et al., 2008). However, our results only showed that the human lipoproteins possess a polarity profile typical for the folded-partially folded proteins and this correlation was rather poor (47% efficiency). We assume that these results can be explained by the relative insensitivity of the method (which assessed the degree of “disorder” for the entire protein) to the presence of local disorder in some protein fragments. Also, our method analyzed the intrinsic polarity of the polypeptide chain of the protein without considering the fact that the lipid has a polarity too, and therefore a specific weight should be added to the final profile of the cargo-protein. To appropriately address these issues, it will be necessary to measure the polarity profile of a carrier protein with and without the lipid, as well as the profile of the lipid transported.

It is pertinent to mention that the polarity index method corresponds to the parallelism scheme known as “master-slaves” (Dijkstra, 1968), so its processing time is 1/n, where “n” is the number of processors in the computer. Therefore, it is possible to use this method to analyze the entire set of proteins and protein regions of a fixed length “n”. For example, if one would aim to analyze all possible protein fragments of the length of 5 amino acids, the number of such fragments would be $20^5 = 3200000$. Assuming that the processing time spent by the method to assess each protein/fragment would be 0.001s, then the time required to analyze all possible fragments would be $3200000 \times 0.001s = 3200s$. However, if the computer would have 10 processors, then the processing time would be just 320s (one tenth part of the computing power on a uni-processor system). Using this approach, it is possible now to process protein fragments of the length of up to 13 amino acids, because the supercomputers are usually formed by more than 4000 processors.

### Table 3. Classification of Lipoproteins

<table>
<thead>
<tr>
<th>Lipoproteins → Lipoproteins</th>
<th>HDL</th>
<th>LDL</th>
<th>VLDL</th>
<th>Chylomicrons</th>
<th>Atherosclerosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL</td>
<td>89</td>
<td>27</td>
<td>35</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>LDL</td>
<td>53</td>
<td>84</td>
<td>47</td>
<td>43</td>
<td>55</td>
</tr>
<tr>
<td>VLDL</td>
<td>21</td>
<td>20</td>
<td>36</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Chylomicrons</td>
<td>21</td>
<td>12</td>
<td>35</td>
<td>86</td>
<td>36</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>11</td>
<td>8</td>
<td>24</td>
<td>21</td>
<td>73</td>
</tr>
</tbody>
</table>

Number of hits (%) for the each pair of Lipoproteins groups (Magrane, 2011), according to polarity index method (Polanco et al., 2012), at the level of efficiency greater than 75%.

### Table 4. Lipoproteins versus Structural profiles

<table>
<thead>
<tr>
<th>Lipoproteins → Structural</th>
<th>Folded</th>
<th>Partially folded</th>
<th>Unfolded</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDL</td>
<td>27</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>LDL</td>
<td>59</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>VLDL</td>
<td>12</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Chylomicrons</td>
<td>15</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>13</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Number of hits (%) for Lipoproteins (Magrane, 2011) in regards to their structural profiles (Oldfield et al., 2005), according to polarity index method (Polanco et al., 2012), at the level of efficiency equal to 100%

### Table 5. Structural profiles versus Lipoproteins

<table>
<thead>
<tr>
<th>Structural → Lipoproteins</th>
<th>HDL</th>
<th>LDL</th>
<th>VLDL</th>
<th>Chylomicrons</th>
<th>Atherosclerosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Folded</td>
<td>37</td>
<td>31</td>
<td>18</td>
<td>29</td>
<td>36</td>
</tr>
<tr>
<td>Partially folded</td>
<td>16</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Unfolded</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

Number of hits (%) for structural profiles (Oldfield et al., 2005) in regards to the Lipoproteins (Magrane, 2011), according to polarity index method (Polanco et al., 2012), at the level of efficiency greater than 75%.
CONCLUSIONS

The Polarity index method is an effective and simple algorithm that can be used as a “front-line filter” for the construction of computational tools for identification and characterization of lipoproteins, as well as for the analysis of protein regions with parallel computing.

Availability

The test-files, and polarity index method program must be requested from the corresponding author (polanco@unam.mx).

Conflict of Interests

We declare that we do not have any financial nor personal interest with other people or organizations that could inappropriately influence (bias) our work.

Author Contributions

Theoretical conception and design: CP. Computational performance: CP. Data analysis: CP, TB, VU, and JACG. Tables: RZ. Results discussion: CP, TB, VU, and JACG.

Acknowledgements

The authors want to thank Concepción Celis Juárez whose suggestions and proof-reading have greatly improved the original manuscript, and we also acknowledge the Computer Science department at Institute de Ciencias Nucleares at the Universidad Nacional Autónoma de México for support, we gratefully acknowledge financial support by the Mexican-French bilateral research grant CONACYT (188689) - ANR (12-IS07-0006).

Figure 1. Distribution of the relative frequencies of the six groups of Lipoproteins. The x-axis represents the 16 polar interactions (Materials & Methods, Metrics).
REFERENCES


Vol. 63       239

REFERENCES
a Ca2+–binding receptor with potential intracellular signaling properties. Eur J Biochem 239: 132–137.


