Analysis of EPO Images after Isoelectric Focusing and Double Blotting

I.Bajla, I.Holländer
Department of High Performance Image Processing
ARC Seibersdorf research GmbH
A-2444, Austria
e-mail: ivan.bajla@arcs.ac.at

G.Gmeiner, and C.Reichel
Doping Control Laboratory
ARC Seibersdorf research GmbH
A-2444, Austria

ABSTRACT

A software system GASEPO1 has been developed as a tool to visualize and analyze doping with recombinant erythropoietin (EPO). Digital images derived from the separation of EPO isoforms by isoelectric focusing (IEF) followed by double blotting and chemiluminescence detection are automatically analysed and evaluated. In the GASEPO1 a novel know-how for image processing and analysis is implemented. In particular, a novel method of cut-off-line calculation has been developed and tested.

KEY WORDS
bioinformatics, doping control, EPO detection, electrophoretic gel image analysis

1 Introduction

Peptide hormones like EPO are widely used in endurance sports like cross country skiing or cycling because of their performance enhancing effect and the difficulties to detect their application. The main goal of EPO doping is the increase of red blood cell production which results in increase of the oxygen transport capacity of the blood. Thus, performance boosts of up to 10 % can be achieved. Since the late eighties, EPO is available on the market in various synthetic or recombinant forms (rEPO). While human EPO, a glycoprotein, is produced in kidneys, rEPO is produced by e.g. chinese hamster ovary cells. With the aid of biotechnology and molecular engineering, production of rEPO for pharmaceutical applications became possible. The possibility to produce rEPO comparably cheaply and its unrestricted availability has catalyzed the use of it as a doping substance.

Conventional doping substances, like anabolic steroids, are detectable by gas-chromatography (GC) and mass-spectrometry (MS). However, these techniques are not capable to provide unique proof of rEPO application because glycoproteins cover a different molecular weight range, much too high to fit into the range of GC-MS.

Previous research has shown that recombinant EPO differs from human EPO in post-translational modifications. This difference manifests itself in different charge ratios of sugar moieties. It was found that for detection of such small differences isoelectric focusing (IEF, using the gradient of pH) is a method of choice. However, due to very low concentrations of endogenous and recombinant EPO among all the other proteins present in human urine the detection of rEPO present in a urine sample is very difficult. F.Lasne proposed in [1] to solve this problem by double-blotting. Such techniques use antibodies for detection and identification of proteins. The process involves the separation of EPO isoforms on a polyacrylamide gel followed by the transfer of the proteins onto a thin membrane (blot). An EPO-specific antibody is incubated on the membrane resulting in a mirror image of the first membrane. Detection of the isoforms is done by a chemiluminescence reaction after incubation of a second antibody, an enzyme catalyst and luminescence reagent.

Figure 1. An EPO gel image example.

After analog or digital imaging a typical pattern of lanes (vertical stripes) is finally generated. As can be seen in Fig. 1, the lanes comprise bands of individual isoforms, which have been separated by pH gradient. When a sample containing rEPO is submitted to IEF, a shift to more basic isoforms is observed compared to endogenous EPO (upper part of the first lane in Fig. 1 with 4–6 bands at different pH positions). When urine with natural uEPO (urine EPO) is submitted to the same process, the spots observed (7–15 bands) are separated.

These bands partially overlap the region of those ones belonging to rEPO (lanes N.3–7 in Fig. 1).
The detection of a positive doping case (i.e. detection of rEPO in presence of endogenous EPO) is based on setting the reference cut-off-line and on calculation of ratio between sum of intensities (profile area) of the lane bands above the cut-off-line and overall intensity sum in the whole lane.

2 GASEPO1 - a software system for the analysis of EPO images

At doping control laboratories worldwide the EPO images are mostly evaluated by a combination of manual approach and using various nonspecialized software packages. Unfortunately these are not tailored to this specific task. The use of different software tools for data interpretation interferes the search for a common basis of gel image interpretation. Up to now no final decision has been made concerning the exact criterion of positivity. To provide a systematical basis for this decision and to contribute to the process of standardization and harmonization in this area, an interdisciplinary team of two departments of ARC Seibersdorf research GmbH has developed a pilot software system GASEPO1 (Gel Analysis System for EPO, Fig. 2).

Due to a number of image degradation and distortion effects occurring in digitized EPO images in practice, several problems of image analysis had to be addressed:

- for noise suppression we have used the geometry-driven diffusion filter which we described in [2],
- we have implemented efficient algorithms for rectification of the geometrical distortions by bicubic spline interpolation that uses a regular grid of knots graphically adjustable by the user,
- for the correction of intensity background inhomogeneity we have developed an algorithm which uses vertical stripes between adjacent lanes without bands,
- a novel method of the calculation of cut-off-line has been proposed, implemented and successfully tested,
- the overall concept of the program system makes use of our previous image analysis system for gel analysis of DNA fragments [3].

3 The conventional and the novel method for cut-off-line calculation

In Fig. 3 a conventionally used method of calculation of the reference cut-off-line ([4]) is outlined. It is based on a combination of using standard graphical tools of universal software packages and manual analysis. A profile of the mean row intensities in the image of a standard rEPO lane is plotted in the right side of Fig. 3. The user searches for the peak (profile local maximum) that corresponds to the most acidic band of the rEPO standard lane denoted as $P_1$. Afterwards the valley $V_1$ between two adjacent maxima in basic direction is found and the distance $\delta$ is measured. The cut-off-line position is finally defined as a position given by $\delta$ distance from $P_1$ peak in direction opposite to the valley $V_1$. This technique assumes, of course, a symmetrical
shape of the profile peak corresponding to the crucial (most acidic) band of the rEPO standard.

For development of a software that would be tailored particularly to analysis of digitized EPO images, it is necessary to implement an automatic method of searching for cut-off-line. In practice, the background of a lane image is not ideally homogeneous, it is corrupted by noise and degraded by various local artefacts. These factors may result in appearance of a number of false local maxima of the profile which have to be separated from the maxima representing the bands. Thus, it is necessary to introduce a flexible threshold to which the local profile maxima are to be compared. The practical problem is how to calculate the value of such a threshold automatically. In Fig. 3 we illustrate this problem by depicting two possible threshold values \( \text{thr}_1, \text{thr}_2 \) which lead to two different positions of the cut-off-line. To design a mathematical algorithm, that could separate the correct maxima from the incorrect ones, is very difficult. The only way how to realize the conventional method of cut-off-line calculation in a program is to leave the decision to the user who can change the value of the threshold interactively. This is evidently a disadvantage of the existing, conventionally used method.

To illustrate the situation when the conventional method is implemented in a program, we analysed 34 images provided from 7 doping control laboratories worldwide. The proper cut-off-line positions (ground true) have been approved by an expert and the threshold necessary for the conventional method was set equal to the average intensity value within the given standard lane. Then we checked whether this value has to be modified in order to find the maximum of relevant peak of the mean profile of the lane. In the given set of EPO images 15 rEPO standard lanes and 80 combined (rEPO+NESP) standard lanes occurred. We have obtained three cases of relative positions of the cut-off-line calculated by the program with the conventional method implemented: (A) the cut-off-line coincides with ground true, (B) the cut-off-line is slightly shifted, even though the threshold \( \text{thr} \) has been set properly, (C) the cut-off-line has unacceptable position and it requires interactive adjusting of the threshold \( \text{thr} \). It should be emphasized that the adjustment had no unique tendency, i.e., for some images a threshold increase was required, while for others its decrease yielded an acceptable result.

For a fixed value of the threshold, there were a number of cases of combined standard lanes in the given EPO image set, when the threshold for the rEPO standard had to be adjusted, while the cut-off-line for the NESP standard was found accurately. The obtained results for both standards are summarized in Table 1.

<table>
<thead>
<tr>
<th>% of cases with</th>
<th>rEPO stand.</th>
<th>NESP stand.</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact position</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>inaccurate position</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>not acceptable position</td>
<td>63</td>
<td>72</td>
</tr>
</tbody>
</table>

To illustrate the situation when the conventional method is implemented in a program, we analysed 34 images provided from 7 doping control laboratories worldwide. The proper cut-off-line positions (ground true) have been approved by an expert and the threshold necessary for the conventional method was set equal to the average intensity value within the given standard lane. Then we checked whether this value has to be modified in order to find the maximum of relevant peak of the mean profile of the lane. In the given set of EPO images 15 rEPO standard lanes and

**Figure 3. Definition of the conventional cut-off-line.**

Based on the above commented computer tests, we have decided to develop a novel, more robust method of calculation of the cut-off-line. It does not use 1D intensity profile and its local maxima, but introduces a concept of partition of the whole lane into two neighboring rectangular block images. In an rEPO standard lane we can assume that the set of bands is located in upper acidic part of the gel. Then we can consider two specific image blocks in the lane image: (i) the upper image block which bounds all bands as tightly as possible, and (ii) the complementary image block in the lane which contains only background. These two image blocks differs considerably in variability of intensities. In ideal case, the former block comprises a harmonic-like varying pattern of band intensities, while the latter one represents just homogeneous background. Therefore the neighboring line separating these two blocks is a good estimate of the cut-off-line. For finding this separating line by an automatical procedure, we propose the following methodology:

1. a set \( \{A_k, B_k\} \) of all possible partitions of the lane image \( I(x, y) \) into neighboring rectangular blocks \( A_k, B_k, k = 1, 2, \ldots, r \) (see Fig. 4) is constructed,
2. we characterize each image block by a measure of intensity variability,
3. for every position of the separating line of two adjacent blocks we define two functions: (i) a function \( \text{topdown} \) given by the values of intensity variability
measure of the blocks \( A_1, A_2, \ldots, A_r \), and (ii) a function bottomup given by the values of intensity variability measure of the blocks \( B_1, B_2, \ldots, B_r \).

4. finally, we define a difference function of these two functions and search for a maximum of this function which represents a pair of adjacent image blocks for which the difference between their intensity variability is the greatest; the position of the maximum determines the cut-off-line.

We have evolved this methodology into the design of a particular algorithm that takes into account specificity of lane EPO images. We describe the steps of this algorithm.

To suppress the image noise a median profile \( P_{med}(x) \) of the intensities of the lane \( I(x,y) \) is first calculated by finding medians of intensity values in each row \( x \) of the lane:

\[
P_{med}(x) = \text{median}\{ I(x,y) : y \in [c,d] \},
\]

(for denotation of the individual parameters see Fig. 4). To smooth nonrelevant maxima we filter the 1D median profile by convolution with a mean filter kernel \( \phi(x) \):

\[
P_{fil}(x) = P_{med}(x) \otimes \phi(x).
\]

The choice of a convenient measure of the intensity variability represents the crucial point of the method. We propose to use the following differential measures characterizing individual image blocks, \( m_t \) for the upper block and \( m_b \) for the lower block:

\[
m_t(A_k) = \frac{1}{a_k - a} \int_a^{a_k} \frac{dP_{fil}(x)}{dx} \, dx,
\]

\[
m_b(B_k) = \frac{1}{b_k - b} \int_b^{b_k} \frac{dP_{fil}(x)}{dx} \, dx.
\]

Finally, the cut-off-line position is defined as the index \( K \) of the maximum difference

\[
\max_k |m_t(A_k) - m_b(B_k)|.
\]

There are several specific problems which have to be solved when implementing the proposed method. We will briefly address them.

First, the range of the image block pairs which are used in practice should be limited. As computer experiments showed, small image blocks \( A_k \), in the extreme case a block just one row high, cannot properly characterize the difference in intensity variability of blocks containing bands. Therefore, it is reasonable to limit the range of the possible height of image blocks included into the system of needed block pairs. For this purpose we denote the average (empirically found) value of the band height within a wide class of EPO images as \( h_{av} \). Let the size of the discrete lane image \( I(i,j) \) is \( (m,n) \). We calculate the average intensity value (threshold) \( I_{av} \) in the whole lane. In the filtered median profile \( P_{fil} \) we search for the first local maximum from the top whose value is greater than \( I_{av} \).

Let the index of this value in the given coordinate system is \( iM \). Thus, the first relevant lane image block \( A_1 \) is defined as the following subimage of the lane:

\[
A_1 = \{ I(i,j) : i = 1, 2, \ldots, (iM - [h_{av}/2]); j = 1, 2, \ldots, n \},
\]

where \( [h_{av}/2] \) is a rounded integer.

Similarly, we define the last lane image block \( B_r \) using the information on the average band height in EPO images:

\[
B_r = \{ I(i,j) : i = m - h_{av}, \ldots, m; j = 1, 2, \ldots, n \}.
\]

Along with rEPO, NESP (Novel Erythropoiesis Stimulating Protein) is also being used as a doping substance with similar effects. It is characterized by the IEF distribution reverse to the rEPO: a set of bands is shifted to the acidic side of the gel (isoforms can be found in the bottom part of the lane). The processing of the NESP standard lane differs from the rEPO standard lane processing in two additional steps: first, before applying the algorithm tailored for the rEPO lane image, we flip up the NESP lane along its horizontal axis, in order to get an auxiliary image in which bands are concentrated in the upper part of the image, characteristic for the rEPO standard. Then the scheme developed for the rEPO standard lane is applied to the resulted geometrically transformed lane image. Finally, the obtained index of the cut-off-line is recalculated to the coordinate system of the initial input lane image.

Combined standard lanes (rEPO and NESP) are often used. For such cases an additional part of the algorithm has been developed:

- For a combined standard lane \( I(i,j) \) with rEPO and NESP standard bands, a 1D median profile \( P_{med}(i) \) is calculated as in the case of pure rEPO standard lane.
• The median profile is then filtered by a mean filter with a large window which is approximately 10% of the lane height \( m \). For the resulted filtered median profile \( P_{fil}(i) \) we search for the global maximum \( \text{Max}_1 \) and the second greatest global maximum \( \text{Max}_2 \).

• A global minimum \( \text{min} \) in the range between the two mentioned global maxima is found. Its index (coordinate) determines the location, where the input lane image is split into two auxiliary subimages, the upper \( L_{e\text{po}} \), and the lower \( L_{n\text{sp}} \). The partial filtered median profiles corresponding to these two subimages are subsets of the initial filtered median profile \( P_{fil}(i) \). We calculate the average intensity value \( I_{av} \) in each of these subimages.

• In the partial filtered median profile for the upper auxiliary lane image we search for the first local maximum from the bottom whose value is greater than \( I_{av} \). The index \( T_{nsp} \) of this maximum serves for the upper limitation of the NESP subimage \( L_{nsp} \) which is defined as follows:

\[
L_{nsp} = \{ I(i, j) : i = T_{nsp} + \lfloor h_{av}/2 \rfloor, \ldots, m - \lfloor h_{av}/2 \rfloor \}.
\]

• In the partial filtered median profile for the lower auxiliary lane image we search for the first local maximum from the top whose value is greater than \( I_{av} \). The index \( B_{e\text{po}} \) of this maximum serves for the lower limitation of the rEPO subimage \( L_{e\text{po}} \) which is defined as follows:

\[
L_{e\text{po}} = \{ I(i, j) : i = 1 + \lfloor h_{av}/2 \rfloor, \ldots, B_{e\text{po}} - \lfloor h_{av}/2 \rfloor \}.
\]

The algorithm tailored for rEPO standard lanes is applied then to the rEPO subimage \( L_{e\text{po}} \).

The algorithm tailored for the NESP standard lanes is applied then to the NESP subimage \( L_{nsp} \).

There are rare cases when degradation of the EPO images by artefacts is so severe (blobs occur in homogeneous parts of lanes) that the assumption on two considerably different image blocks, is violated. In such cases the user has to cut out the crucial region from the affected lane.

In Fig. 5 we illustrate cut-off-line positions calculated by the conventional method (white dashed lines) and the novel method (black solid lines).

4 Conclusions

• We have described a pilot program system GASEPO1 which has been developed under MATLAB\textsuperscript{TM} in collaboration of two departments of the ARC Seibersdorf research GmbH. The system has been extensively tested in real doping control practice.

• Its basic functions are: loading images, contrast setting, selection of a region of interest (ROI), rectification of geometrical distortions, lane analysis, generating report, saving the results. In Fig. 6 an illustration of the typical report generated by the GASEPO1 system is displayed.

• In the system an innovative method for calculation of the reference cut-off-line is implemented. A proposal for Austrian patent covering this method is pending. The scrupulous testing of the system on images from seven doping control laboratories showed robust behaviour of the novel cut-off-line procedure.

• To make the analysis for the user more instructive and transparent, the system GASEPO1 is provided by a number of additional user-friendly tools and utilities, which include 3D visualization. In Fig. 7 an illustration of this tool for cut-off-line representation in 3D is displayed.
Figure 6. Illustration of cut-off-lines for rEPO (upper) and NESP (lower) in the typical report of image analysis generated in the GASEPO1.

Figure 7. 3D plot illustration of the cut-off-line.

References


