Available online at www.sciencedirect.com
SCIENCE DIRECT
Polymer 46 (2005) 2791–2798
polymer
www.elsevier.com/locate/polymer

Stochastic molecular descriptors for polymers. 2. Spherical truncation of electrostatic interactions on entropy based polymers 3D-QSAR

Humberto González-Díaz[a,b,*], Liane Saíz-Urra[b], Reinaldo Molina[b,c], Eugenio Uriarte[a]

[a]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela 15706, Spain*
[b]*Chemical Bioactives Center, Central University of "Las Villas" 54830, Cuba*
[c]*Universität Rostock, FB Chemie, Albert-Einstein-Str. 3a, D 18059 Rostock, Germany*

Received 1 October 2004; received in revised form 12 January 2005; accepted 12 January 2005

Abstract

The spherical truncation of electrostatic field with different functions break down long-range interactions at a given cutoff distance (r_{off}) resulting in short-range ones. Consequently, a Markov Chain model may approach to the entropies of spatial distribution of charges within the polymer backbone. These entropies can be used to predict polymers properties [González-Díaz H, Molina RR, Uriarte E. Polymer 2004; 45: 3845 [53]]. Herein, we explore the effect of abrupt, shifting, force shifting, and switching truncation functions on QSAR models classifying 26 proteins with different function: lysozymes, dihydrofolate reductases, and alcohol dehydrogenases. Almost all methods have shown overall accuracies higher than 85% instead of 80.8% for models based on physicochemical parameters. The present result points to a acceptable robustness of the Markov model for different truncation schemes and r_{off} values. The results of best accuracy 92.3% with abrupt truncation coincides with our recent communication [Bioorg Med Chem Lett 2004; 14: 4691–4695]. Nonetheless, the simpler model with three variables and high accuracy (88%) was derived with a shifting function and $r_{off} = 10$ Å.

Keywords: Polymers 3D structure; Markov chains; Electrostatics

1. Introduction

Polymers, in special biopolymers such as proteins and nucleic acids, are highly charged molecules for which an accurate treatment of the long-range electrostatic interactions is very important in molecular dynamics (MD) approximations to polymers structure [1–3]. Nonbonded pairwise interactions between atoms or groups are usually truncated at a specific cutoff distance (r_{off}) to reduce the number of interactions and thereby the required computational time for the simulation [4–7]. That is to say, it is common practice to neglect long-range interactions beyond the cutoff distance. Such spherical cutoffs can be implemented in different ways, depending on whether the distance is calculated between the interacting atoms (atom-based) or between two groups of atoms (group-based). Furthermore, the interaction energy or force can be *truncated* abruptly at the cutoff distance, or some kind of smoothing scheme can be applied, either on the whole range 0, r, r_{off} (a *shift*), or over a limited region r_{on}, r, r_{off} (a *switch*). The shifting or switching function $S(r)$, as described below for the different cases, multiplies Coulombs law to give the effective form of the electrostatic interaction used in the calculations as recently reviewed in a seminar work by Norberg and Nilsson [8]. The importance of long-range interactions in biomolecular systems has been reviewed by other authors too [9–10]. However, the uses of spherical truncation approaches have been restrained to the field MD mainly ignoring possible applications on the developments of proteins 3D structure molecular descriptors. The search of novel molecular descriptors in the range of small-to-medium sized molecules in order to seek quantitative-structure-activity-relationships (QSAR) [11] constitutes nowadays a widely covered field with more than 1 000 molecular descriptors introduced [12,13]. Some of these indices have encountered interested applications on the field

* Corresponding author.
E-mail address: humbertogd@vodafone.es (H. González-Díaz).

0032-3861/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2005.01.066

of polymers indeed [14–17]. By the contrary, the search for theoretic approaches reaching to new molecular descriptors for biopolymers have began more after in spite of an early (pioneer) work of Flory in 1953 on the radius of gyration [17]. More recently appeared other approaches which are potential sources, define, or apply in some extent polymer descriptors, such as the Arteca's mean over crossing number, the Randic's band average widths, the sequence-order-coupling numbers, α-helix-propensity descriptors, Emini Surface Index, the SDA (sum of cosines of dihedral angles), Kyle–Dolittle hydrophbicity, and the I3 index [18–27]. In any case, the search of molecular descriptors for biopolymers structure facing to QSAR studies is an emerging area.

On the other hand, Markov Chain (MC) models are well-known tools for characterizing biomolecules structure. MC models have been used for analyzing biological sequence data and they have been used to find new genes from the open reading frames. Another use of these models is data-based searching and multiple sequence alignment of protein families and protein domains. Protein turn types and sub-cellular locations have been successfully predicted. Hubbard and Park used amino acid sequence-based hidden MC models to predict secondary structures. In this sense, Krogh et al. have also proposed a hidden MC model architecture. In addition, Markov's stochastic process has been used for protein folding recognition. This approach can also be used for the prediction of protein signal sequences. Another seminar works can be found related to the application of MC theory to Proteomics and Bioinformatics. Chou applied MC models to predict beta turns and their types, and the prediction of protein cleavage sites by HIV protease [28–42]. Anyhow, have not been reported many works on Markov models for the generation of molecular descriptors encoding proteins 3D structure facing to QSAR.

In this connection, our group has introduced elsewhere a physically meaningful Markov model (Markovian Chemicals In Silico Design: MARCH-INSIDE) encoding molecular backbones information. It allowed us introducing matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties. Specifically, the entropy like molecular descriptors has demonstrated flexibility in many different problems such as: anticoccidial, flukicidal, and anticancer drugs design as well as prediction of drug-induced agranulocytocis. In the field of polymers the method has been applied to model the interaction between drugs and HIV-RNA, and predicting proteins and virus activity as well. In a very recent communication we reported the use of MARCH-INSIDE to encode polymers structures, e. g. proteins, in QSAR studies with abrupt truncation of the electrostatic field [43–54].

Consequently, we will describe herein a number of studies that have focused on the advantages or disadvantages of different truncation methods for long-range electrostatic interactions on proteins 3D-QSAR using MC molecular descriptors.

2. Methods

Consider a representation for a polymer, e.g. protein, described as a static model, which considers a spatial distribution of pseudo monomers, e.g. aminoacids, with 3D coordinates (x_i, y_i, z_i) coinciding with those for a reference atom in the polymer, e.g. the Cα for an aminoacid. In this case, every pair of monomers in the polymer backbone (i, j) present a pairwise electrostatic interaction with energy E_{ij}. The electrostatic charge (q_i) will be considered to be equal to the electronic charge of the monomer. In the case of aminoacids we can consider those reported by Collantes and Dunn [55]. As a result, it is then easy to deal with the problem of the propagation of the effect of all monomer–monomer (aminoacid–aminoacid) pairwise electrostatic interactions within the polymer (protein) backbone. All of these E_{ij} may be determined using Coulomb's formula. If we then arrange all these interaction energies in a matrix and normalize the values dividing by row sums, we obtain a stochastic matrix ${}^1\Pi(x, y, z, q)$. This step makes it possible to study the propagation of the electrostatic interactions within the protein backbone as a MC. In doing so, the elements of ${}^1\Pi(x, y, z, q)$ may be considered as the probabilities (${}^1p_{ij}$) with which the monomer (aminoacid) i presents a truncated electrostatic interaction of energy E_{ij}, with the monomer (aminoacid) j placed at a distance r_{ij} [50,54,55]:

$$ {}^1pij = \frac{S_{ij}^2(r)E_{ij}}{\sum_{m=1}^{\delta+1} S_{im}^2(r)E_{im}} = \frac{S_{ij}^2(r)q_iq_j/r_{ij}^2}{\sum_{k=1}^{\delta+1} S_{im}^2(r)q_iq_m/r_{im}^2} \quad (1a) $$

where it is straightforward to realise that the use of a probabilistic formulation determines the simplification of the q_j charges. That is to say, it is equivalent to use energy (Eq. (1a)) or an electrostatic potential (ϕ_j) interpretation (Eq. (1b)).

$$ {}^1pij = \frac{S_{ij}^2(r)q_j/r_{ij}^2}{\sum_{m=1}^{\delta+1} S_{ik}^2(r)q_k/r_{im}^2} = \frac{S_{ij}^2(r)\varphi_{ij}}{\sum_{k=1}^{\delta+1} S_{im}^2(r)\varphi_{im}} \quad (1b) $$

In Eqs. (1a) and (1b) the sum consider all the δ monomers (aminoacids) that have a spherical truncated interaction with the monomer (aminoacid) i. In other words, in this study the electrostatic field was transformed from a continuous field to a discrete field, making a direct MC matrix codification possible. As described above, in the introduction section, $S_{ij}(r)$ is the truncation function, which has different formulation in dependence of the method used, i.e. a shifting function like in Eq. (2a) or a force-shifting function like in Eq. (2b). Alternatively, we will explore also a switching function like in Eq. (2c) [8,56,57]:

$$ S_{ij}(r) = \begin{cases} \left(1 - \left(\frac{r}{r_{off}}\right)^2\right)^2, & r \le r_{off} \\ 0, & r > r_{off} \end{cases} \quad (2a) $$

$$S_{ij}(r) = \begin{cases} \left(1 - \frac{r}{r_{off}}\right)^2, & r \le r_{off} \\ 0, & r > r_{off} \end{cases} \quad (2b)$$

$$S_{ij}(r) = \begin{cases} 1, & r \le r_{on} \\ \dfrac{(r_{off}^2 - r^2)(r_{off}^2 + 2r^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3}, & r_{on} < r < r_{off} \\ 0, & r \ge r_{off} \end{cases} \quad (2c)$$

The main approximation here is undoubtedly to consider that a spherical truncated electrostatic interaction may propagate throughout space to other aminoacids in the protein as a MC. The present approach neglects long-range Coulomb interactions (dotted arrow) in the stochastic matrix but conversely to classic truncation methods allows for them in a step-by-step fashion (solid arrows), as shown in Fig. 1.

The spatial dependence of the model becomes clear on inspection of Fig. 1. Due to truncation restrictions, the monomer (aminoacid) a_0 (with charge q_0 and coordinates x_0, y_0, z_0) is only able to interact with the monomer (aminoacid) a_1 by means of direct interaction. The effect of this interaction can only affect monomer (aminoacid) a_2 in a subsequent propagation of the interaction a_1-a_2 and so on. It is clearer here identifying the parameter of the MC as the topologic distance or number of steps (k) one interaction needs to propagate from one aminoacid to other instead of the time, which is the more classic MC parameter. However, one should note that this number of elemental steps (k) one truncated interaction uses to propagate throughout space are given at corresponding discrete time intervals ($\Delta t_k = t_{k+1} - t_k = k$) like in almost MC applications [46,51].

We will consider that the absolute probabilities ($^{A}p_k(j)$) with which these long-range interactions rise to a specific aminoacid j after k steps are the elements of the vectors $^{k}\Phi(x, y, z, q)$ derived in a Markovian manner using the so-called Chapman–Kolmogorov equations [48,49,52–54]:

$$^{k}\Phi(x,y,z) = {}^{0}\Phi(x,y,z,q)\,{}^{k}\Pi(x,y,z,q) = {}^{0}\Phi(x,y,z,q)[{}^{1}\Pi(x,y,z,q)]^{k} \quad (3)$$

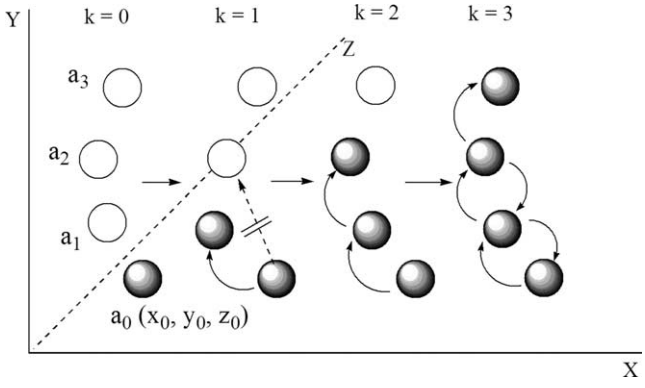

Fig. 1. Illustrative step-by-step propagation of the interaction within the polymer backbone.

where $^{0}\Phi(x, y, z, q)$ is the vector in which elements ($^{A}p_0(j)$) are the initial absolute probabilities with which any aminoacid j participates in an electrostatic interaction. After this simple approximation, calculating the spectrum of entropies Θ_k with which the effect of the electrostatic interactions propagate until a distance k throughout the polymer (protein) backbone is relatively straightforward:

$$\Theta_k(G) = -k_B \sum_{j \subset G} {}^{A}p_k(j) \log {}^{A}p_k(j) \quad (4)$$

where G define a specific group of monomers (aminoacids) having a defined condition, e.g. styrene monomers, polar aminoacids, aromatic aminoacids, nucleosides forming and hydrogen bond. That is to say, the entropy may be calculated for the polymer as a whole or as a local parameter [45].

3. Results and discussion

Truncation approaches have been applied in different MD studies to a broad range of polymers and cutoff distances [58]. A comparison between the Ewald and the switching function techniques was performed for a zwitterionic pentapeptide in aqueous solution by Smith and Pettitt at cutoff distances of 9–10 Å [59]. The protein HIV-1 protease was simulated by York et al. for 300 ps in its crystal environment using a residue-based approach with a cutoff of 9.0 Å [60]. An extensive study by Loncharich and Brooks, focused on carboxymyoglobin and analyzed six methods of truncating the long-range interactions in MD simulations including values of cutoff of 14 Å [61].

In order to determine the effect of using different long-range electrostatic field truncation approaches in polymers 3D-QSAR we have developed a linear discriminant analysis to find a QSAR for 26 proteins. These proteins in spite of similar folding have three different biological activities namely: Lysozymes (L), dihydrofolate reductases (DR), and alcohol dehydrogenases (AD). Briefly, all the truncation method were used on the lookout for significant QSAR models, see Fig. 2 for overall accuracy, but one can note some specific points:

(a) Shifting function: all the QSAR models presented high values of the canonical regression coefficient $R_c > 0.80$. Interestingly, the overall accuracy of the models increases with r_{off} from 77% ($r_{off} = 7$ Å) up to 88% ($r_{off} = 13$ Å), see Table 1. Specifically, the atom-based approach ASH using the shifting function [56] and a cutoff of 12.0 Å performs fairly well. These results coincide with Kitson et al. findings, who found good results on modelling the protein *Streptomyces griseus* protease A even at very large cutoff distances of 25 Å [62].

(b) Force shifting function: presented a similar behaviour than the shifting function in connection to the variation

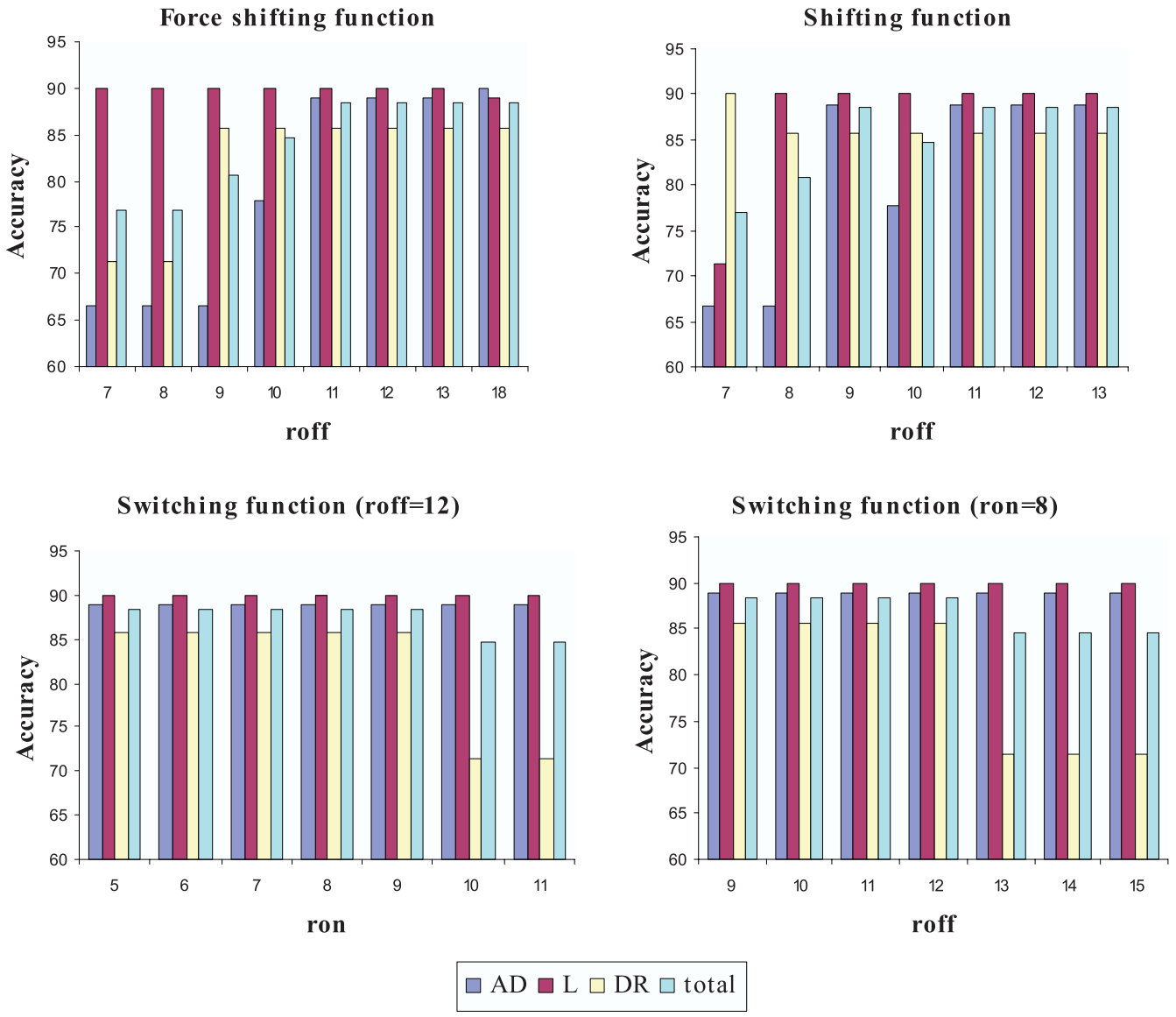

Fig. 2. Bar graphics for total and group accuracy with different truncations approach.

of the overall accuracy with respect to r. Although, for this truncation scheme the change the changes in overall accuracy (77–81, 81–85, and 85–88%) have taken place at smaller $r_{off}=9$, $r_{off}=10$, and $r_{on}=11$ values with respect to similar changes for the shifting function, which appeared at $r_{off}=8$, $r_{off}=9$, and $r_{off}=10$. This difference of 1 Å may be related to the dependence of the functions with r, see Eqs. (2a) and (2b). It is notable that the atom-based approach with the force-shift method [56,57] and a short cutoff of 8.0 Å (AFSHS) presented less accuracy than its middle an and longer range analogues AFSH and AFSHL, see Table 1.

(c) Switching function: we developed two experiments expanding the domain of the switching function both inward ($r_{off}=12$ and r_{on} variable) and outward ($r_{on}=8$ and r_{off} variable). In both cases the QSAR presented good accuracy particularly ASW a very useful truncation approach [56,57], previously reported, performed very well, see Table 1.

(d) In general all the QSAR models have performed better with lysozymes than with the other proteins specifically at cutoff near to 10.0 Å. Similar results where reported by Saito, who recommend optimum cutoff of 10.0 Å modelling human lysozyme [63], see Table 1.

(e) As a tendency, the shifting, and switching perform better at higher cutoff coinciding with Steinbach and Brooks results, which argued that applying a cutoff of 12.0 Å or longer is more important than choosing a specific truncation method [64], see Table 1.

(f) Abrupt truncation function: this approach presented the higher accuracy 92% at $r_{off}=50\%$ of the van der Waal distance among atoms, in coincidence with our previous

Table 1
Summary of the results for different truncation approaches

Truncation scheme: shifting function								
r_{off}	Scheme name	AD%	L%	DR%	Total %	λ	F	R_c
7[a]		67	71	90	77	0.20	6.26	0.83
8[a]		67	90	86	81	0.19	6.54	0.83
9[a]		78	90	86	85	0.21	5.96	0.83
10[a]		89	90	86	88	0.19	6.45	0.82
11[a]		89	90	86	88	0.21	5.88	0.83
12[a]	ASH	89	90	86	88	0.21	5.78	0.83
13[a]		89	90	86	88	0.22	5.69	0.83
Truncation scheme: force shifting function								
r_{off}	Scheme name	AD%	L%	DR%	Total %	λ	F	R_c
7[a]		67	90	71	77	0.18	6.65	0.84
8[a]	AFSHS	67	90	71	77	0.20	6.31	0.83
9[a]		67	90	86	81	0.19	6.60	0.83
10[a]		78	90	86	85	0.19	6.50	0.83
11[a]		89	90	86	88	0.19	6.40	0.82
12[a]	AFSH	89	90	86	88	0.21	5.91	0.83
13[a]		89	90	86	88	0.21	5.84	0.83
18[a]	AFSHL	90	89	86	88	0.22	5.57	0.82
Truncation scheme: switching function with $r_{off}=12$ and variable r_{on}								
r_{on}	Scheme name	AD%	L%	DR%	Total %	λ	F	R_c
5[a]		89	90	86	88	0.22	5.62	0.83
6[a]		89	90	86	88	0.22	5.60	0.83
7[a]		89	90	86	88	0.22	5.57	0.83
8[a]	ASW	89	90	86	88	0.23	5.50	0.83
9[a]		89	90	86	88	0.23	5.48	0.83
10[b]		89	90	71	85	0.25	6.98	0.82
11[b]		89	90	71	85	0.25	6.98	0.82
Truncation scheme: switching function with $r_{on}=8$ and variable r_{off}								
r_{off}	Scheme name	AD%	L%	DR%	Total %	λ	F	R_c
9[a]		89	90	86	88	0.22	5.74	0.83
10[a]		89	90	86	88	0.22	5.72	0.83
11[a]		89	90	86	88	0.22	5.62	0.83
12[a]	ASW	89	90	86	88	0.23	5.52	0.83
13[b]		89	90	71	85	0.25	6.98	0.82
14[b]		89	90	71	85	0.25	6.98	0.82
15[b]		89	90	71	85	0.25	6.98	0.82
Truncation scheme: abrupt truncation								
R_{off} %	Scheme name	AD%	L%	DR%	Total %	λ	F	R_c
50[b]		89	90	100	92	0.09	11.1	0.90
60[a]		89	90	71	85	0.23	5.48	0.84
70[a]		89	90	71	85	0.22	5.62	0.84
80[a]		89	90	71	85	0.22	5.56	0.84
90[a]		89	90	71	85	0.22	5.68	0.84

[a] Models with three variables.
[b] Models with four variables.

reports [54] and models implemented in almost docking and MD software [65]. However, the method accuracy abruptly decays to 85% for every $r_{off}>60\%$, see Table 1.

All the QSAR models studied have three equations as a consequence of a three group (L, AD, DR) LDA analysis. Nevertheless, we are going to report only the better QSAR models found herein taking into consideration the higher accuracy, less number of variables, and simpler cutoffs procedure. In this sense, the model derived with abrupt truncation function presented the higher accuracy 92% at $r_{on}=50\%$, and the simpler truncation approach (abrupt truncation) but it is not the simpler one having

four variables [54]:

$$L = 1.17\Theta_5(c) + 49.1\Theta_0(m) - 26.5\Theta_3(s) + 74.3\Theta_0(T) - 753.7$$

$$AD = 1.22\Theta_5(c) + 53.6\Theta_0(m) - 28.9\Theta_3(s) + 82.8\Theta_0(T) - 930.0$$

$$DR = 1.09\Theta_5(c) + 48.1\Theta_0(m) - 24.6\Theta_3(s) + 75.9\Theta_0(T) - 781.0 \quad (5a)$$

where c, m, s, and T are specific groups or collections of aminoacids placed at protein core (c), middle region (m), surface (s), or in every place (T) [54] Conversely, the model for shifting function with r_{on} of only 10 Å have a very good accuracy too and have only three variables, see Table 1:

$$L = 0.78\Theta_1(c) + 9.732\Theta_0(m) + 28.96\Theta_0(T) - 349.5$$

$$AD = 1.29\Theta_1(c) + 10.49\Theta_0(m) + 32.75\Theta_0(T) - 274.8$$

$$DR = 2.72\Theta_1(c) + 9.49\Theta_0(m) + 31.28\Theta_0(T) - 349.5 \quad (5b)$$

These entire models herein studied confirm the high potentialities of the methods describing polymer structure as a function of molecular descriptors based on truncation approaches [66] and the concept of entropy [67,68].

In closing, we would like to discuss the effect of the dielectric constant and the homogeneity for charge distribution. First, making an analysis of the dielectric function in the equation of probability (1a) three different cases can be detected for the dielectric function:

1. The case in which the dielectric function is constant for the whole polymer/media system $\varepsilon = \text{cte}$. In this case the dielectric constant is extracted from the sum as a common factor and then simplified in such a way that the calculated probability does not depends on ε an the equations (6a) reduce to the particular case (1a):

$${}^1pij = \frac{S_{ij}^2(r)q_iq_j/\varepsilon r_{ij}^2}{\sum_{k=1}^{\delta+1} S_{im}^2(r)q_iq_m/\varepsilon r_{im}^2} = \frac{(q_i/\varepsilon)S_{ij}^2(r)q_j/r_{ij}^2}{(q_i/\varepsilon)\sum_{k=1}^{\delta+1} S_{im}^2(r)q_m/r_{im}^2} = \frac{S_{ij}^2(r)\varphi_{ij}}{\sum_{k=1}^{\delta+1} S_{im}^2(r)\varphi_{im}} \quad (6a)$$

2. The case in which the media dielectric function depends on one or more macroscopic parameters x, like the temperature (T). As an example we can refer to the temperature dependence of chloromethane dielectric with T [69]:

$$\varepsilon_{(T)} = 12.6 - 0.061 \times (T + 20) + 0.0005 \times (T + 10)^2 \quad (6b)$$

Other example is the T and pressure (P) dependence of simple gases dielectric function $\varepsilon(T,P)$ correction with respect to the dielectric constant $\varepsilon(20\ ^\circ\text{C},P)$ at 20 °C and 1 atm [69]:

$$\varepsilon_{(T,P)} = 1 + \frac{(\varepsilon_{(20\ ^\circ\text{C},1\ \text{atm})} - 1)P}{760(1 + 0.003411(T - 20))} \quad (6c)$$

In these cases, due to the macroscopic characteristic of the parameters (T and P) the dielectric function has the same value in all the system. Therefore, the dielectric function can be simplified too and the calculated probability does not depend on ε as in the Eqs. (1a) and (6a).

3. The case of media dielectric function depends on local parameters as the distance among the neighboring amino-acids [70]. In this case it is impossible to simplify the term and one has to substitute ε for its expression in function of distance. In these cases the values of the molecular descriptors depend on the values of function $\varepsilon_{(rij)}$ selected and the QSAR equation have to be optimized with respect to this value.

$${}^1pij = \frac{S_{ij}^2(r)q_i/\varepsilon_{(r_{ij})}r_{ij}^2}{\sum_{k=1}^{\delta+1} S_{im(r)}^2 q_i/\varepsilon_{(r_{ij})}r_{im}^2} \quad (6d)$$

Finally, the distribution and degree of charges sites over a chain could influence notably the conditions for spherical truncation. In proteins with semi-uniform charge distribution (see next scheme) the distance between the charges may be considered approximately equal (r_{su}), semi-uniform distance (see Fig. 3). If we truncate the electrostatic field within the protein at $r_{off1} < r_{su}$, then $S_{ij}(r) = 0$ and the elements of the matrix ${}^1\Pi$ vanished (${}^1p_{ij} = 0$), then one can not calculate any molecular descriptors. On the other hand, if we truncate the electrostatic field within the protein at $r_{off2} \geq r_{su}$, $S_{ij}(r) = f(r)$ and $p_{ij} > 0$ for all aminoacids pair-wise interaction. Consequently, for large 'semi-uniformly

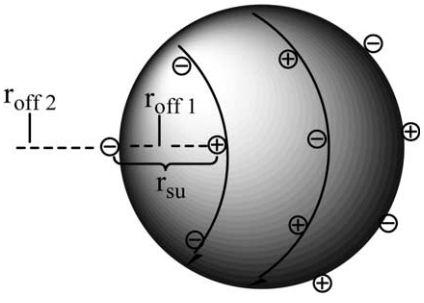

Fig. 3. Illustration of the effect of semi-uniform distribution and degree of charges chain sites in.

distributed' highly charged globular proteins calculations have to be done at $r_{off2} \geq r_{su}$ and may becomes time consuming. In this case computers are needed more potent than in other cases in order to perform the calculation in a reasonably lower time.

4. Conclusions

In summary, we would draw four main conclusions from this study:

1. This work introduces by the first time stochastic molecular descriptors for polymers QSAR studies considering truncated electrostatic interactions within the 3D-backbone.
2. In order to save time calculation of these molecular descriptors may be simplified with different truncation approaches.
3. Studies should be carry out to determines at which conditions of number of variables, truncation approach, and cutoffs one can find the best QSAR models.
4. In this specific case, we shown how the model applies to the study of proteins function with the best results in terms of accuracy and simplicity for abrupt truncation and shifting function approaches, respectively.

Acknowledgements

González-Díaz H and Uriarte, E. would like to express his sincere gratitude to the Xunta of Galicia for grant PR405A2001/65-0, which was used to purchase the Statistica 6.0 software. The editor Prof. J. E. Mark and one unknown referee are acknowledged by kind attention and very interesting recommendations.

References

[1] McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. Cambridge, UK: Cambridge University Press; 1987.
[2] Saenger W. Principles of nucleic acid structure. New York: Springer-Verlag; 1988.
[3] Navarro E, Fenude E, Celda B. Biopolymers 2004;73:229.
[4] Sagui C, Darden TA. Annu Rev Biophys Biomol Struct 1999;28:155.
[5] Guenot J, Kollman PA. J Comput Chem 1993;14:295–311.
[6] Harvey SC. Proteins 1989;5:78–92.
[7] Esteve V, Blondelle S, Celda B, Perez-Paya E. Biopolymers 2001;59:467.
[8] Norberg J, Nilsson L. Acc Chem Res 2002;35:465–72.
[9] Berendsen HJC. Electrostatic interactions. Computer simulations of biomolecular systems: theoretical and experimental applications. Leiden, The Netherlands: ESCOM; 1993. pp. 161–181.
[10] Smith PE, Van Gunsteren WF. Methods for the evaluation of long range electrostatic forces in computer simulations of molecular systems. Computer simulations of biomolecular systems: theoretical and experimental applications. Leiden, The Netherlands: ESCOM; 1993. pp. 182–212.
[11] Kubinyi H, Taylor J, Ramdsen C. Quantitative drug design. In: Hansch C, editor. Comprehensive medicinal chemistry, vol. 4. Oxford: Pergamon; 1990. p. 589–643.
[12] Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim: Wiley VCH; 2000.
[13] Randic M. In: Schleyer PVR, editor. Encyclopedia of computational chemistry, vol. 5. New York: Wiley; 1998. p. 3018.
[14] González MP, Morales AH. J Comput Aid Mol Des 2003;10:665.
[15] Morales AH, González MP, Rieumont JB. Polymer 2004;45:2045.
[16] González MP, Dias LC, Morales AH. Polymer 2004;15:5353.
[17] González MP, Morales AH, Molina R. Polymer 2004;45:2773.
[18] Flory PJ. Principles of polymer chemistry. Itaha: Cornell University Press; 1953.
[19] Arteca GA. J Chem Inf Comput Sci 1999;39:550.
[20] Arteca GA, Mezey PG. J Mol Graph 1990;8:66.
[21] Randić M, Vračko M, Nandy A, Basak SC. J Chem Inf Comput Sci 2000;40:1235.
[22] Randić M, Balaban AT. J Chem Inf Comput Sci 2003;43:532.
[23] Hua S, Sun Z. Bioinformatics 2001;17:721.
[24] Cai YD, Lina SL. BBA 2003;1648:127.
[25] Lejon T, Strom BM, Svensen SJ. J Pept Sci 2002;7:74.
[26] Gutman I, Rosenfield VR. Theor Chim Acta 1996;93:191.
[27] Estrada E. Chem Phys Lett 2000;319:713.
[28] Vorodovsky M, Koonin EV, Rudd KE. Trends Biochem Sci 1994;19:309.
[29] Vorodovsky M, Macininch JD, Koonin EV, Rudd KE, Médigue C, Danchin A. Nucl Acids Res 1995;23:3554.
[30] Krogh A, Brown M, Mian IS, Sjeander K, Haussler D. J Mol Biol 1994;235:1501.
[31] Chou KC. Biopolymers 1997;42:837.
[32] Yuan Z. FEBS Lett 1999;451:23.
[33] Hua S, Sun Z. Bioinformatics 2001;17:721.
[34] Hubbard TJ, Park J. Protein: Struct Funct Genet 1995;23:398.
[35] Krogh A, Brown M, Mian IS, Sjeander K, Haussler D. J Mol Biol 1994;235:1501.
[36] Di Francesco V, Munson PJ, Garnier J. Bioinformatics 1999;15:131.
[37] Chou KC. Curr Protein Pept Sci 2002;3:615.
[38] Chou KC. Peptides 2001;22:1973.
[39] Chou KC. Anal Biochem 2000;286:1.
[40] Chou KC. J Biol Chem 1993;268:16938.
[41] Chou KC. Anal Biochem 1996;233:1.
[42] Chou KC, Zhang CT. J Protein Chem 1993;12:709.
[43] González-Díaz H, Olazábal E, Castañedo N, Hernádez SI, Morales A. J Mol Mod 2002;8:237.
[44] González-Díaz H, Hernández SI, Uriarte E, Santana L. Comput Biol Chem 2003;27:217.
[45] González-Díaz H, Marrero Y, Hernández I, Bastida I, Tenorio I, Nasco O, et al. Chem Res Toxicol 2003;16:1318.
[46] González-Díaz H, Gia O, Uriarte E, Hernádez I, Ramos R, Chaviano M, et al. J Mol Mod 2003;9:395.
[47] González-Díaz H, Ramos de AR, Uriarte E. Online J Bioinf 2002;1:83.
[48] González-Díaz H, Ramos de AR, Molina R. Bull Math Biol 2003;65:991.
[49] González-Díaz H, Ramos de AR, Molina R. Bioinformatics 2003;19:2079.
[50] Ramos de AR, González-Díaz H, Molina RR, Uriarte E. Proteins: Struct Funct Bioinf 2004;56:715.
[51] González-Díaz H, Bastida I, Castañedo N, Nasco O, Olazabal E, Morales A, et al. Bull Math Biol 2004;66:1285.
[52] Ramos de AR, González-Díaz H, Molina R, González MP, Uriarte E. Bioorg Med Chem 2004;12:4815.
[53] González-Díaz H, Molina RR, Uriarte E. Polymer 2004;45:3845.
[54] González-Díaz H, Molina RR, Uriarte E. Bioorg Med Chem Lett 2004;14:4691.
[55] Collantes ER, Dunn WJ. J Med Chem 1995;38:2705.

[56] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. J Comp Chem 1983;4:187.
[57] Brooks III CL, Pettitt BM, Karplus M. J Chem Phys 1985;83:5897.
[58] Darden TA, Toukmaji LG. Pedersen J Chim Phys 1997;94:1346.
[59] Smith PE, Pettitt BM. J Chem Phys 1991;95:8430.
[60] York DM, Darden TA, Pedersen LG. J Chem Phys 1993;99:8345.
[61] Loncharich RJ, Brooks BR. Proteins 1989;6:32.
[62] Kitson DH, Avbelj F, Moult J, Nguyen DT, Mertz JE, Hadzi D, et al. Proc Natl Acad Sci USA 1993;90:8920.
[63] Saito M. J Chem Phys 1994;101:4055.
[64] Steinbach PJ, Brooks BR. J Comp Chem 1994;15:667.
[65] Navarro E, Fenude E, Celda B. Biopolymers 2002;64:198.
[66] Monleon D, Celda B. Biopolymers 2003;70:212.
[67] González-Moa M, Terán MC, Mosquera RA. Int J Quantum Chem 2002;86:7.
[68] Lorenzo L, Mosquera RA. J Comput Chem 2003;24:707.
[69] Weast RC, editor. CRC handbook of chemistry and physics. Cleveland, OH: The Chemical Rubber Co.; 1971.
[70] Norberg J, Nilsson L. Biophys J 2000;79:1537.