Available online at www.sciencedirect.com
ScienceDirect
Polymer 48 (2007) 5788–5801
polymer
www.elsevier.com/locate/polymer

Prediction of fibrinogen adsorption for biodegradable polymers: Integration of molecular dynamics and surrogate modeling

Anna V. Gubskaya [a,*], Vladyslav Kholodovych [b], Doyle Knight [c], Joachim Kohn [a,*], William J. Welsh [b]

[a] New Jersey Center for Biomaterials, Rutgers, The State University of New Jersey, 145 Bevier Road, Piscataway, NJ 08854, United States
[b] Department of Pharmacology, University of Medicine and Dentistry of New Jersey (UMDNJ)-Robert Wood Johnson Medical School (RWJMS), Piscataway, NJ 08854, United States
[c] Department of Mechanical and Aerospace Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States

Received 2 February 2007; received in revised form 2 July 2007; accepted 2 July 2007
Available online 12 July 2007

Abstract

This work is a part of a series of publications devoted to the development of surrogate (semi-empirical) models for the prediction of fibrinogen adsorption onto polymer surfaces. Since fibrinogen is one of the key proteins involved in platelet activation and the formation of thrombosis, the modeling of fibrinogen adsorption on the surface of blood-contacting medical devices is of high theoretical and practical significance. We report here, for the first time, on the incorporation of three-dimensional structures of polymers obtained from atomistic simulations into conventional mesoscopic-scale calculations. Low energy conformations derived from molecular dynamics simulations for 45 representatives of a combinatorial library of polyarylates were used in an improved modeling procedure (referred to as ''3D surrogate model'') instead of simplistic two-dimensional representations of polymer structures, which were used in several previous models (collectively referred to as ''2D surrogate models''). In the framework of this 3D model we created 12 model sets of polymers to account for their chirality, conformational diversity and the structural influence of a solvent. For each polymer set, three-dimensional molecular descriptors were generated and then ranked with respect to the experimental fibrinogen adsorption data by means of a Monte Carlo decision tree. The most significant descriptors identified by decision tree and the experimental dataset were utilized to predict fibrinogen adsorption using an artificial neural network (ANN). The best prediction achieved by the 3D surrogate model demonstrated a noticeable improvement in the predictive quality as compared to the previously used 2D model (as evidenced by the increase in the average Pearson correlation coefficient from 0.54 ± 0.12 to 0.67 ± 0.13). The predictive quality of the 3D surrogate model compares favorably with the best results previously reported for extended 2D model that combines an ANN with partial least squares (PLS) regression and principal component (PC) analysis. The significance of the newly developed 3D model is that it allows high accuracy prediction of fibrinogen adsorption without the need for experimentally-derived descriptors and it has better predictive quality than the original 2D surrogate model due to utilization of realistic polymer representations.

Keywords: Biodegradable polyarylates; Prediction of fibrinogen adsorption; Computer-aided design

1. Introduction

1.1. Prediction of polymer material properties

Rapid developments in the field of combinatorial design of new biologically relevant polymeric materials have led to an inevitable growth in the number of potential biomaterial candidates [1]. This trend makes it practically impossible to experimentally evaluate the important polymer material properties in large combinatorial libraries, and, consequently a search for new methods that allow prediction of polymer material properties becomes paramount. Two major computational approaches – QSPR (quantitative structure–property relationship) models and machine learning – have been successfully used to this end. In both methods the chemical

* Corresponding authors. Tel.: +1 732 445 3888; fax: +1 732 445 5006.
E-mail address: kohn@biology.rutgers.edu (J. Kohn).

0032-3861/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2007.07.007

structure of a polymer is usually represented by a set of descriptors, which are drawn from experimental properties or computed *ab initio*. Landrum et al. [2] have shown that a careful choice of descriptors allows efficient predictive models to be built. These authors identified ten descriptors that captured the relevant physics and chemistry of the problem and successfully predicted the molecular weight of polymers using machine learning algorithms such as hierarchical clustering, *k*-nearest neighbors (KNN) algorithm, decision trees, support vector machines (SVMs) and bag classifiers [2]. Recently Duce et al. [3] suggested that in the case of polymers the need for molecular descriptors restricts the applicability of predictive models, and proposed a recursive neural network (RecNN) that directly incorporates the structured representation of the polymers based on the two-dimensional (2D) graph of the repeating unit treated as a rooted chemical tree [3].

QSPR models have been widely used to predict the glass transition temperature (T_g) of amorphous polymers, the fundamental property that affects heat capacity, coefficient of thermal expansion and viscosity of a polymer. The most referenced predictive model of T_g was produced by Bicerano [4] and subsequently numerous researchers have attempted to achieve the same level of accuracy in their predictions [4–7]. Beyond the problem of the accurate estimation of T_g, other polymer properties relevant to their end-use performance have been the focus of developers of predictive models. Among these, the prediction of three-dimensional polymer structures with exceptional mechanical properties [8], qualitative and quantitative predictions of composition of multicomponent bioglasses [9], the computer-aided design of polymer blends [10], and polymer–receptor interactions in biosensors [11] are worth mentioning. Surprisingly, the prediction of protein adsorption onto polymer surfaces, cell attachment, and cell proliferation (collectively referred to in this article as the ''bioresponse'') have thus so far not been extensively modeled.

1.2. Prediction of fibrinogen adsorption using surrogate modeling and combined models

Protein adsorption onto polymer surfaces plays a crucial role in biomaterials design [12]. Adsorption of fibrinogen on polymer surfaces is especially important within the context of blood-contacting medical implants, due to fibrinogen's role in platelet activation and thrombus formation. Fibrinogen is the most abundant of the adhesive proteins present in plasma and can therefore be taken as a marker for the tendency of a polymer surface to support the attachment and subsequent activation of platelets. The predictive modeling of fibrinogen adsorption on polymer surfaces is of value to a wide range of biomaterials scientists and biomedical engineers.

We have recently reported on the design of surrogate (semi-empirical) models [13–15] for the prediction of cellular response to and fibrinogen adsorption on the surfaces of tyrosine-derived polyarylates synthesized in Kohn's research group [16–18]. In these studies, the polyarylates served as a ''test case'' for the validation of the fibrinogen adsorption models – the methodology is quite general and can be readily adopted to other polymer libraries. Surrogate modeling of fibrinogen adsorption was performed in two stages. First, 104 [14] or 862 [15] two-dimensional (2D) molecular descriptors were calculated for each polymer based on its molecular structure and combined with two experimentally measured descriptors: glass transition temperature (T_g) and air–water contact angle (θ). Then the significance of each descriptor with respect to the experimental fibrinogen adsorption data was determined by means of a decision tree methodology. In the second stage, the most significant descriptors in conjunction with measured fibrinogen adsorption data were used to build an artificial neural network (ANN) model. This model [14] required experimentally-derived descriptors (i.e., T_g and θ) as well as 104 computed descriptors to achieve sufficiently accurate predictions that allowed one to identify polymers with the highest and lowest fibrinogen adsorption. Since for most practical biomedical implant applications either low or high protein adsorption is required, the ability to identify these ''extreme'' polymers is a significant advance that allows avoiding experimental evaluation of non-optimal polymer candidates.

While the above modeling approach provided some time savings, the need to include experimentally-derived descriptors in the model was a limitation. In principle, it should be possible to predict fibrinogen adsorption on virtual polymer surfaces in the absence of any experimentally-derived descriptors. Toward that goal, the same group of authors [19] developed an extended surrogate 2D model that eliminated the experimentally determined T_g and θ from the set of 106 descriptors. To maintain the accuracy of the model, several aspects of the ANN were combined with partial least squares (PLS) regression and principal component analysis (PCA) – well-known techniques in conventional QSAR approaches [20]. On the one hand, PLS regression and PC analysis can be employed independently for surrogate modeling of fibrinogen adsorption. Alternatively, they both can be used to select precisely the most significant combinations of descriptor variables, and, hence, to account for synergy between apparently uncorrelated parameters. Augmenting the ANN by the PLS-extracted descriptor set provided better representation of the nonlinear relationship between the structural features of the polymer and fibrinogen adsorption on the polymer surface [19].

1.3. The rationale for combining molecular dynamics (MD) and surrogate modeling

Both methodologies described above require calculation and input of molecular descriptors for each polymer. Smith et al. designed two successful surrogate models [14,19] to predict fibrinogen adsorption onto polymer surfaces in which 2D chemical descriptors were chosen to quantify the structure of polymers at the molecular level. This class of descriptors has proven to be fairly successful in prediction of fibrinogen adsorption by means of QSAR analysis in spite of the fact that these descriptors do not take into account the conformational specificity of each polymer, which is known to play a crucial role in interactions between the proteins and the

surfaces [12]. Although X-ray or NMR nuclear Overhauser effect structural experiments generally serve as the best source of reliable 3D information, these methods are too labor intensive to be useful for the creation of 3D descriptors of polymer libraries containing large numbers of individual polymers. As an alternative, we suggest that molecular dynamics (MD) simulations represent the most attractive choice: the subtle details of molecular motion, interactions and structure can be extracted readily from computer simulations performed under realistic conditions of temperature and pressure. MD simulations of polymers have a relatively long and well-established history and, in the present work, are used to determine 3D conformations of 45 representatives selected from the polyarylate library. This allows expansion of the database of input parameters for the ANN model by including more complex three-dimensional molecular descriptors that describe 3D polymer structure and possibly reflect molecular level characteristics of polymer surfaces. The objective of the present work is three-fold, namely: (1) generation of an ANN model based upon the 3D structure of the polymer obtained from MD simulations, (2) investigation of the influence of different levels of 3D organization within the framework of this model on its accuracy, and (3) comparative assessment of the results obtained with those previously reported by Smith et al. [14] in order to estimate the predictive quality of both 2D and 3D-based modeling designs and identify further directions for their improvement.

2. Materials/methods/background

2.1. Combinatorial library design

The combinatorial library of polyarylates was created using a parallel synthesis approach developed in Kohn's laboratory [16–18]. Starting from the pioneering work of Smith and co-authors [13], it has become an ideal dataset for calibrating and testing computational models for prediction of bioresponse [14,15,19,20]. The library comprises 112 structurally related alternating co-polymers (see Fig. 1). It was synthesized using eight commercially available diacids and 14 tyrosine-based diphenols. Each repeat unit of a polymer consists of a diacid (DA) and a diphenol (DP) component. The structure of these building blocks was varied at Y and R, respectively. The number of methyl groups on the backbone of the DP component is also variable ($n = 1$ or 2, Fig. 1). The naming convention employed for polyarylates is based on the names of their acid and alcohol precursors (see Ref. [14] for detailed description). The names of the 45 polymers utilized in the present work are given in Table 1.

2.2. Experimental methods

The experimental part consists of the preparation of polymer films and the immunofluorescence assay procedure for the measurements of fibrinogen adsorption in a rapid screening format. All results are reported in detail in [21]. The polymers were

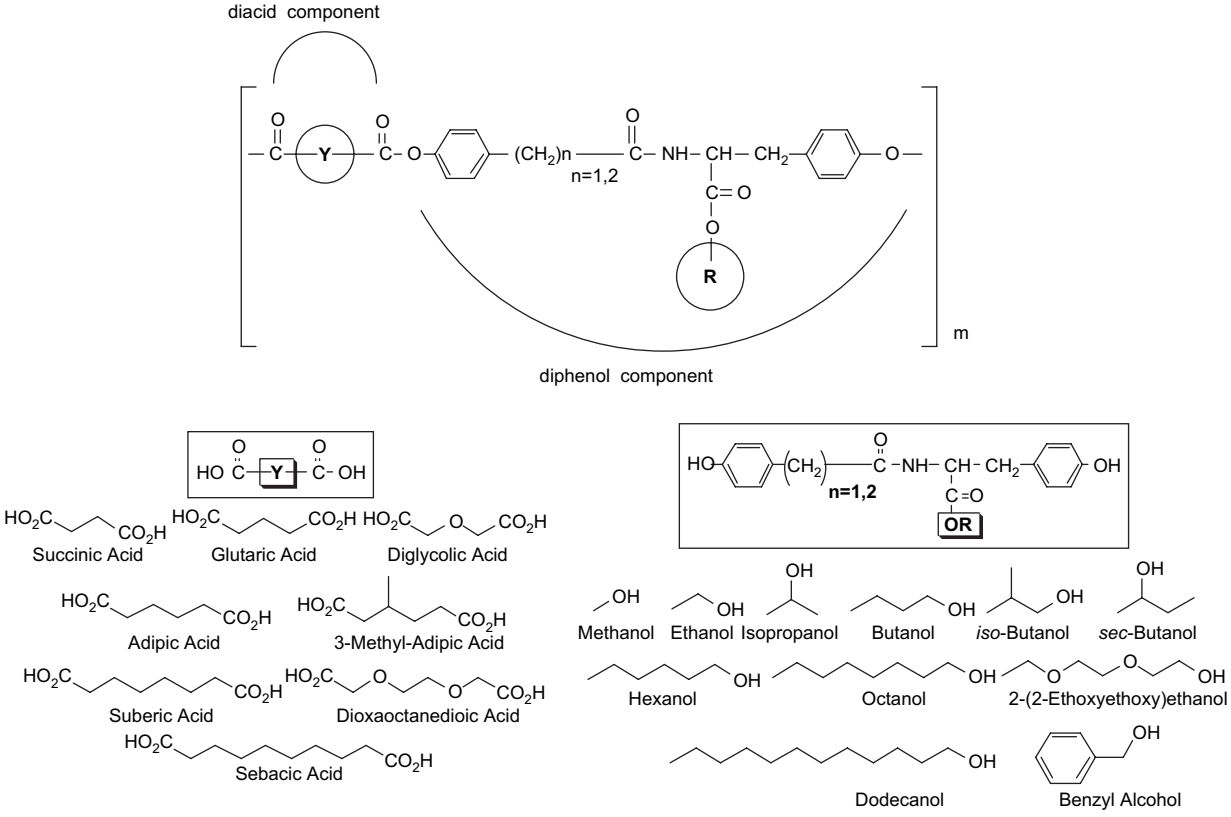

Fig. 1. Combinatorial library of 112 polyarylates. Symbol Y stands for eight diacids (shown on the left side) and symbol R stands for 14 tyrosine-derived diphenols (shown on the right side).

Table 1
Polymer numbering scheme: polymers are placed in the order of increasing experimentally measured fibrinogen adsorption

No.	Pendent	Diacid	No.	Pendent	Diacid
1	DTiB	Sebacate	24	DTB	Glutarate
2	HTH	Sebacate	25	HTE	Adipate
3	DTO	Glutarate	26	DTsB*	Adipate
4	DTO	Sebacate	27	DTM	Methyl adipate*
5	HTH	Adipate	28	DTB	Adipate
6	HTH	Suberate	29	DTB	Succinate
7	DTO	Adipate	30	DTE	Adipate
8	DTBn	Sebacate	31	DTH	Succinate
9	DTO	Suberate	32	DTsB*	Glutarate
10	DTH	Adipate	33	DTBn	Methyl adipate*
11	DTiB	Adipate	34	DTM	Suberate
12	DTH	Suberate	35	DTBn	Adipate
13	DTBn	Suberate	36	DTH	Diglycolate
14	DTH	Methyl adipate*	37	DTO	Diglycolate
15	DTH	Glutarate	38	DTM	Adipate
16	DTM	Sebacate	39	DTiP	Methyl adipate*
17	DTB	Suberate	40	HTE	Methyl adipate*
18	HTE	Suberate	41	DTE	Glutarate
19	DTsB*	Suberate	42	DTsB*	Methyl adipate*
20	DTiB	Succinate	43	DTE	Methyl adipate*
21	DTiP	Adipate	44	HTE	Succinate
22	DTO	Succinate	45	DTB	Diglycolate
23	DTB	Methyl adipate*			

The '*' symbol indicates the presence of more than one chiral center in the polymer repeat unit.

prepared using a solvent casting procedure that can be briefly described as follows. After dissolving polymers in 5% (w/v) methylene chloride, the polymer solutions were filtered through 0.45 μm PTFE filters (Whatman Inc., Clifton, NJ). Then, the polymer solutions were placed on the plates with individual polypropylene microtiter wells and dried in a temperature-controlled oven at 50 °C for 3 h to evaporate methylene chloride and to produce thin and smooth polymer films inside the wells [21]. The immunofluorescence assay protocol includes two incubation steps followed by fluorescence measurements performed using a Spectra Max Gemini fluorescence reader (Molecular Devices, Sunnyvale, CA). A 25 μL aliquot of fibrinogen diluted with phosphate-buffered saline (PBS) to 3 mg/μL concentration was incubated into polyarylate-filled microwells on a polypropylene plate for 1.5 h at 37 °C. After rinsing eight times with PBS, the wells were incubated again with 1% (w/v) bovine serum albumin (BSA) in PBS for 0.5 h at 37 °C to block nonspecific antibody binding. Subsequently, the rinsing procedure was repeated three times and a background fluorescence measurement was taken. Fluorescently labeled antibodies were prepared [21] and allowed to bind to fibrinogen adsorbed onto polymer surfaces for 1.5 h at 37 °C. Then the final fluorescence measurements were carried out with human FA to non-coated propylene used as an internal control.

2.3. Computational procedure

The computational tasks of this project can be summarized in a stepwise manner as follows: generation of 3D structures of the polymers, structure minimization, MD simulations, calculation of 3D descriptors, ranking descriptors with respect to experimental fibrinogen adsorption data, and prediction of fibrinogen adsorption. This protocol includes the sequential use of commercial simulation packages such as MacroModel v. 8.5 (Schrödinger [22]) and DRAGON (Milano Chemometrics and QSAR Research Group [23]) as well a custom modified C5 decision tree [24] and ANN [14] algorithms. A schematic representation of the entire computational procedure is shown in Fig. 2. Such a sequential utilization of different kinds of software (i.e., combination of micro- and meso-scale simulations) usually requires taking into consideration numerous details starting from the compatibility of input–output file formats and ending up with a size of each polymer model compound (e.g., the DRAGON software cannot accommodate a molecular system which exceeds 300 atoms).

Chirality of polymers is an important issue that was also addressed in the present study. According to the definition (conditions of parallel synthesis), L-tyrosine-derived polyarylates have a primary chiral center on the carbon atom next to the amine group. This center always gives rise to L optical isomers. From the list of polymers in Table 1 it is evident that 11 out of 45 selected polyarylates exhibit a second chiral center located in a pendent chain (DTsB) or/and in a diacid component (methyl adipate). From the modeling point of view the presence of two chiral centers requires taking into consideration and generation of all possible enantiomers thus implying additional complexity to the modeling task. Information regarding chirality was requested from manufacturers and it was concluded that the actual polymers are likely represented as 50/50 racemic mixture of L and R isomers with respect to these secondary chiral centers. To simplify modeling of the representation of the chiral species, two separate sets containing exclusively L and R isomers with respect to the secondary chiral

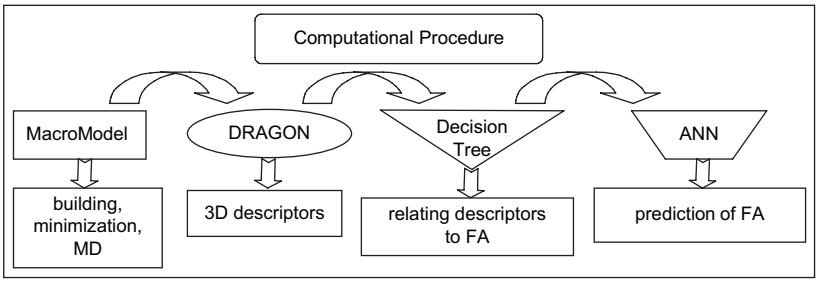

Fig. 2. Schematic representation of computational procedure.

centers were built to account for the influence of chirality in our fibrinogen adsorption predictive models. Thus, the set of 45 polymers employed in this study was constructed in such a way that all polymers in the set are represented only by L optical isomers with respect to the primary chiral center in the backbone and 11 polymers are represented by both L and R enantiomers (see Table 1 for naming convention).

2.4. MD simulations

MD simulations were performed in stage I of this work on the polymers from the polyarylate library for which data on fibrinogen adsorption were previously obtained. To our knowledge this is the first time a MD-based conformational study was performed for the subset of a combinatorial library consisting of 45 structurally related biodegradable polymers. The MacroModel (version 8.5) commercial software [22] with the generalized Born/surface area (GB/SA) implicit solvent model [25] was employed throughout this stage.

Tyrosine-derived polyarylates comprise structural elements that can be found in peptides, carboxylic acids, hydrocarbons and esters. The OPLS-all atom force field was chosen for the present MD simulations [26–28]. This force field is one of the best empirical force fields available for condensed-phase simulations of peptides and it has also been proven to yield comparable results for a wide variety of organic systems. Conveniently, when the MacroModel performs energy calculations, the program checks the ability of the force field in use to mimic the real potential energy of the investigated molecules. This check was performed for all 45 polymers and it indicated only high quality force field parameters.

The initial configurations for MD simulations of amorphous polymer models were constructed as follows. First, a repeat unit for each of the selected polymers was built and its energy was minimized with an all atom representation and no constraints applied. Next, polymer chains containing 3–4 repeat units (to satisfy a system size requirement of 300 atoms) were built with head-to-tail connection. In addition, several model structures of 20 repeat units length were constructed for comparison. The stepwise chain construction scheme, similar to that employed in the conventional rotational isomeric state (RIS) model, was employed to generate the initial structures. The energy of all polymer chains was also minimized in vacuum and in the presence of implicit solvent (water) by means of the Polak-Ribiere conjugate gradient (PRCG) scheme. Convergence using the PRCG with a threshold of 0.05 was achieved after 3000–5000 iterations. Constant temperature (NVT) MD simulations were performed at 298 K maintained by a Nose-Hoover thermostat in vacuum and in the presence of implicit solvent (water) with the integration time step of 1 fs. The cutoff distances employed in vacuum simulations were 7 Å for non-bonded and 12 Å for electrostatic interactions. Extended cutoffs of 8 Å and 20 Å, respectively, for non-bonded and electrostatic interactions were used for simulations in implicit water. Polymer structures were equilibrated for 0.3–0.4 ns until full convergence of the energy. The total simulation time was 1 ns where the latter 0.6 ns was the production time used to collect statistics. To ensure that at the end of the production time the trajectories converged into similar final configurations and that the most significant descriptors derived from such configurations are reproducible, several test simulations were performed using different initial structures generated for the selected subset of polymers. Relative RMSD of atomic positions in final configurations were less than 0.4, 0.7 and 0.9 Å for heavy atoms of a backbone, for heavy atoms of entire molecule and for all atoms of the polymer, respectively.

2.5. Three-dimensional (3D) molecular descriptors

During the past decade it has become apparent that 3D molecular structure must be incorporated in both QSPR and machine learning applications through inclusion of geometrical information and/or physicochemical atomic properties. These so-called 3D descriptors are usually calculated based on either crystallographic coordinates or molecular geometry obtained by computational chemistry. The 3D geometrical representation of a molecule is encoded in RDF (radial distribution function) descriptors, 3D-MoRSE (3D-molecule representation of structures based on electron diffraction) descriptors and WHIM (weighted holistic invariant molecular) descriptors [29]. Requirements for 3D descriptors include their independence of the size of a molecule, uniqueness regarding 3D arrangement of the atoms and invariance with respect to translation and rotation of the molecule. Topological molecular indices represent a special subset of 3D descriptors and are derived from the structural graph of a molecule by using the geometric distances between atoms instead of the topological distances (e.g., Randic molecular profiles). Developments in combinatorial chemistry and high-throughput screening approaches initiated a search for the descriptors that would incorporate as much chemical and structural information as possible and would be suitable for rapid screening of large populations of compounds. The novel molecular descriptors called GETAWAY (geometry, topology and atom-weights assembly) which account for both the geometrical and the topological information weighted by chemical information encoded in selected atomic weightings (unit weights, mass, polarizability, electronegativity) were recently devised [30,31]. They have already shown good predictive capability in physicochemical property modeling and appear to be suitable for similarity/diversity analysis of large chemical databases [31]. In the present work all types of 3D descriptors mentioned above (total 721) were calculated by the DRAGON software [23] for the polymer conformations obtained from local minimization and MD simulations.

2.6. Decision tree approach

Decision tree analysis was employed in stage II of our modeling procedure after the 3D descriptors were generated for each polymer. Decision trees (DTs) are commonly used for the description, classification and generalization of data [32]. The DTs that are involved in descriptor analysis classify data points by starting at the top of the tree (root node) and moving down the tree by creating a hierarchy of descriptor values on an

"if-then-else" basis at each branch point until the terminal (leaf) node is reached. In these top-down constructions the data are recursively divided into subsets based upon the best classifying descriptors at each level. The C5 DT algorithm [24] employed in the present study evaluates the significance of each descriptor with respect to the set of experimental fibrinogen adsorption data using the concept of information gain introduced by Shannon [33,34]. To take into account the experimental uncertainty, the conventional C5 routine was augmented by a Monte Carlo (MC) procedure. A series of 10 000 computer-based pseudo-experiments were generated where the measured values of fibrinogen adsorption were randomly perturbed within a normal distribution defined by the experimental mean and standard deviation. A single most relevant descriptor (i.e., the descriptor with the highest information gain) was obtained from each pseudo experiment. Descriptor results from all MC iterations were summarized into a histogram and the three descriptors with the highest counts in this histogram were taken as input variables for the ANN.

2.7. Artificial neural networks

In stage III of the present study we utilized the most significant descriptors in conjunction with the experimental dataset to build an ANN model. ANNs are machine learning routines that perform nonlinear multidimensional vector mappings [35–37]. Only recently, conventional ANNs have been used to model polymer systems that represent complex classes of highly nonlinear materials. There are several parameters (weights) that need to be defined for ANN. This problem can be handled by augmenting ANNs with genetic algorithms (GA) and Markov chains to enhance the accuracy of prediction [35]. It became a common practice to estimate an accuracy of the model on new data by splitting the dataset randomly into two halves before model building. One half, the training set, is used to train the model while the remaining half, the validation set, is used to validate the model. The ANN in this study is represented by a standard feed–forward network in which every layer is connected to the previous layer. In particular, it employs a two-layer perceptron with two hidden neurons utilizing a sigmoid function with an inverse length scale parameter (κ) equal to 0.1. All input variables were scaled to the unit interval. It was previously shown (see Ref. [14] also for more mathematical details of this ANN model) that the predictive capability of this ANN is essentially independent of the choice of intrinsic variables such as the number of input variables in the net, the number of hidden neurons and the values of κ parameter. The optimization is achieved by means of GADO, a GA specifically designed for optimization domains [14].

To include the experimental variation explicitly in the model, a Monte Carlo approach was used in the generation of the final predicted values. The procedure is similar to that described in the Section 2.5. It was found that a sequence of 400 pseudo-experiments was sufficient to achieve convergence of the final predicted values of the average root-mean-square percentage error and the average correlation coefficient. As noted previously [14], we confirmed that the results of prediction are not entirely independent of the choice of training set. We modified the previous version of ANN by incorporating a MC procedure that generated random multiple splitting of the entire polymer set. Repeating the sequence data split-model training–validation process multiple times and collecting statistics made it possible to improve the estimation of the predictive accuracy of the current ANN model.

It is important, however, to address here the issue of the possible occurrence of a feature selection bias in the used procedure by referring to the articles by Xiong et al. [38] and Ambroise et al. [39] that provide a useful description of potential pitfalls associated with this phenomenon. In particular, Xiong et al. [38] developed a rule for cancer detection based upon gene expression. First, the authors performed a feature selection (i.e., determined the relevant genes) using the entire experimental data set. This is analogous to our use of the entire experimental dataset for fibrinogen adsorption in the decision tree to identify the most relevant descriptors. Second, Xiong et al. developed a rule for detecting cancer based upon the selected genes using 95% of the experimental data and then tested the rule on the remaining 5% of the data. In contrast, we developed a model (analogous to the rule proposed in [38]) using 50% of the experimental data and then tested the model on the remaining 50% of the experimental data. Although there may be some feature extraction bias in our model, it is certainly minimized by the fact that only 50% of the experimental data were used to train the model. The current approach to feature extraction is exactly the same as used by Smith et al. [14,19], which employed descriptors based on 2D polymer structure. The improvement of the accuracy of the present predictive model for fibrinogen adsorption (see Section 3.3 below) is exclusively due to the incorporation of the more relevant 3D descriptors derived from realistic polymer structures.

3. Results and discussion

3.1. Results of molecular mechanics and MD simulations

A noticeable difference in polymer conformations was observed after initial molecular mechanics energy minimization with dramatic variations depending on the chemical composition and chirality of the polymer. Selected examples of minimized structures are shown in Figs. 3 and 4. The difference between enantiomers can be clearly seen for L,L and L,R conformers of poly(DTsB methyl adipate) minimized in vacuum (Fig. 3) and implicit water (not shown). For some polymers helix-like conformers are apparent where the helical shape is defined exclusively by bond and torsion angles and is not stabilized by intramolecular hydrogen bonding. This type of conformers was observed mainly for the polyarylates with a relatively long aliphatic pendent chain and diacid component. In the case of poly(DTE dodecandiaoate) the helical-type structure was observed only in the presence of implicit water.

In our MD simulation we focused mainly on obtaining the low energy relaxed conformation of each polymer rather than calculating average thermodynamic properties. As a relevant example of such properties, the various energy contributions

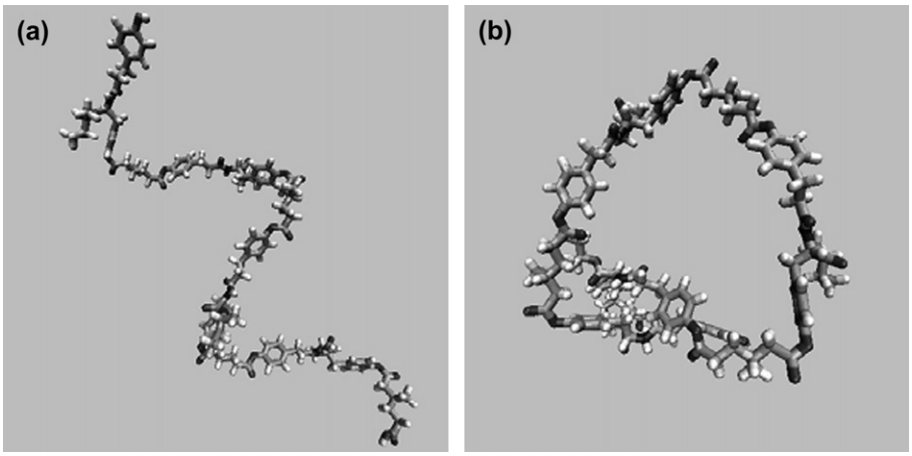

Fig. 3. Poly(DTsB methyl adipate) minimized in vacuum. (a) L,L Enantiomer and (b) L,R enantiomer, where the first and second letter define optical isomerism with respect to the chiral centers on the pendent and diacid, respectively.

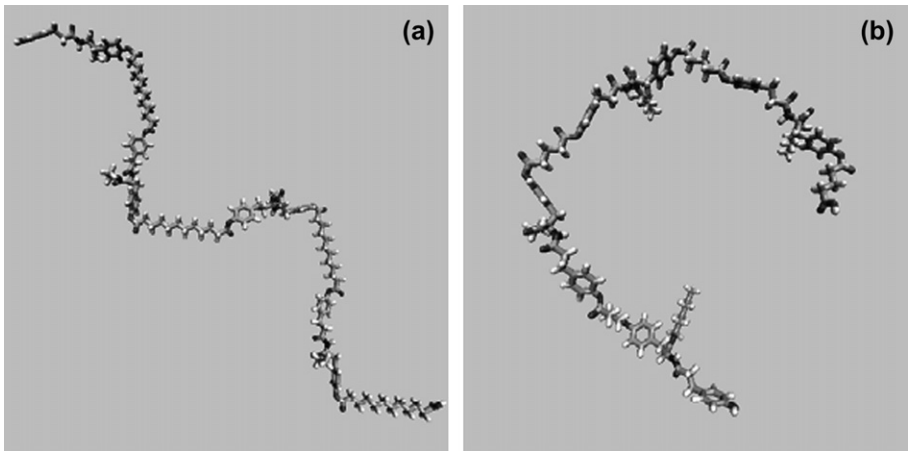

Fig. 4. (a) Poly(DTE dodecandioate) and (b) poly(DTH glutarate) minimized in implicit water.

into the total average potential energy are summarized in Table 2 for two selected polyarylates, which differ only in the length of the aliphatic diacid component. For compared molecules, values of corresponding contributions from bond stretch, angle bend, torsion and solvation terms to the total interaction potential are approximately the same (or change insignificantly). In contrast, VDW and especially electrostatic interactions are responsible for the major difference in the total potential energy between these two polymers. The importance of the electrostatic interactions will also reveal itself in the results of our descriptor analysis (see Sections 3.2 and 3.3).

MD simulations carried out both in vacuum and in implicit water indicated that a folded coil- or globular-like conformational pattern is prevalent for all polymers. It is possible that such a uniform conformational pattern is a result of omitting specific solvent effects (e.g., hydrogen bonding with water molecules) due to utilization of an implicit solvation model [40].

Formation of a globular type conformation did not depend on the length of the polymer chain. The key torsion angles responsible for the folding of the polymer chain in tetramers were comparable with those recorded in 20 monomers long structures (Fig. 5). Some interesting packing trends were observed within a set of 45 polyarylates (see for example Fig. 6). They include a tendency of short pendent chains to be packed inside the globule while the long and especially aliphatic pendants tend to appear on the surface. Additionally, the polymer segments represented by aliphatic hydrocarbons fairly often demonstrate mutual alignment due to hydrophobic interactions. The

Table 2
Various contributions to the total potential energy (kJ/mol) of poly(DTE glutarate) and poly(DTE dodecandioate) simulated in vacuum and implicit aqueous environment

Average energy[a]	DTE glutarate (v)	DTE glutarate (w)	DTE dodecandioate (v)	DTE dodecandioate (w)
Bond stretch	25.05	29.58	19.85	23.13
Angle bend	450.04	466.26	471.60	477.73
Torsion	208.57	201.68	222.12	218.00
VDW	50.55	39.80	17.23	−4.52
Electrostatic	−790.42	−5.88	−590.74	−4.12
Solvation	0.00	−406.21	0.00	−308.97
U total	−56.21	325.23	140.06	401.25

[a] Average potential energy is scaled to 300 K.

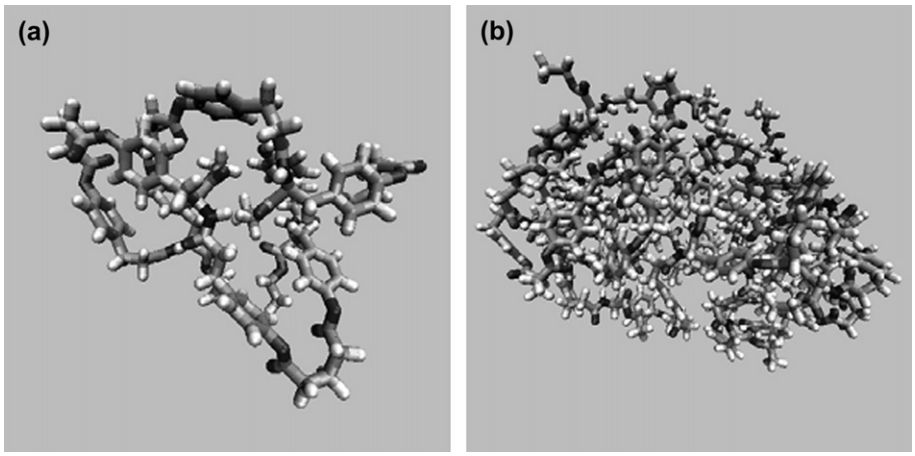

Fig. 5. Final configurations of 0.7 ns MD run in vacuum. (a) Tetramer of poly(DTE glutarate) and (b) chain of 20 monomers of poly(DTE glutarate).

presence of both face-to-face (parallel or π-stacking) and edge-to-face (perpendicular) interaction modes between aromatic rings was also evident. The edge-to-face aromatic interactions are essential in such biologically important processes as protein folding and ligand-receptor interactions [41]. Their role in polymer folding as well as the detailed analysis of the variety of conformational states and interactions found in L-tyrosine-derived polyarylates will be published elsewhere.

To investigate in more detail the sensitivity of our modeling approach to the ''level'' of 3D organization we combined all simulated polymer structures into three large groups that included minimized initial structures, final configurations from MD simulations and minimized final configurations from MD runs. The latter appears to be a convenient approximation of the average polymer structure. In these three conformational groups (or ''levels'') the noticeable structural difference between unfolded initial structures and folded globular-like conformations is juxtaposed with those represented by subtle rearrangements of functional groups during the final minimization of already folded polymer chains. Expansion of this structural classification onto the polymer sets which were comprised from L and R optical isomers and taking into account simulation conditions (vacuum versus implicit water) allowed us to generate 12 polymer sets of 45 polymers each and, therefore, to account for a level of 3D organization, chirality and effect of solvent in our next modeling developments.

3.2. Decision tree descriptors result

To generate an input for the DT, a total of 721 3D molecular descriptors (3 89 340 total for all datasets) were calculated for each polymer in the each of 12 sets. The best descriptors identified by the MCDT for each polymer set are summarized in Table 3 and their definitions are given in Table 4.

One can see that from all possible 3D descriptors introduced in Section 2.5 only representatives of three descriptor groups, namely geometrical, 3D-MoRSE and GETAWAY were found to be significant with respect to fibrinogen adsorption onto polymer surfaces. If the definitions of geometrical descriptors are self-explanatory, the physicochemical meaning encoded by the descriptors from the latter two groups is not always obvious and requires more detailed explanation. The 3D-MoRSE descriptors are derived by mathematical transformations of electron diffraction patterns. This general molecular transformation represents the scattering in various directions by an ensemble of spherical scatters (atoms). Its mathematical

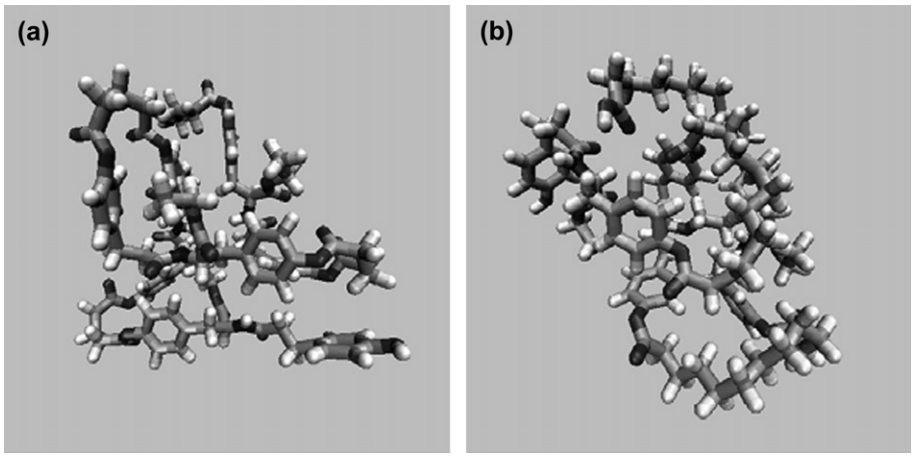

Fig. 6. Final configurations of 1 ns MD simulations in implicit water. (a) Tetramer of poly(DTE glutarate) and (b) tetramer of poly(DTE dodecandioate).

Table 3
The most significant descriptors identified by MCDT for each of 12 sets of polyarylates described in Section 3.1

Lev.	Vacuum		Water	
	L	R	L	R
Minimization				
1	R_5^+(m)	R_5^+(m)	Mor6(m)	Mor28(m)
2	R_6^+(u)	Mor32(u)	R3(u)	Mor19(e)
3	Mor32(u)	Mor3(p)	R_7^+(e)	R_3(u)
MD simulations				
1	Mor19(p)	Mor19(p)	$HATS_3$(m)	$HATS_3$(m)
2	G(N···O)	H_0(e)	$HATS_1$(m)	$HATS_1$(m)
3	G(O···O)	G(O···O)	G(O···O)	G(O···O)
Minimization after MD				
1	R_4(e)	R_4(e)	$HATS_1$(m)	$HATS_1$(m)
2	G(N···O)	Mor7(v)	G(O···O)	G(O···O)
3	Mor7(v)	H_0(e)	R_3(u)	Mor14(m)

definition is given by the expression for the intensity of scattered radiation, which includes the measured scattering angle and the form factors represented by atomic properties such as atomic mass, partial atomic charge, residual atomic electronegativities and atomic polarizabilities [42]. These descriptors allow the size-independent representation of the 3D structures of polymers by a fixed number of variables (i.e., 32 3D-MoRSE signals) where bioactivity is implicitly tied to electronic properties of specific sites of polymers [42,43]. In contrast GETAWAY descriptors are not obtained from experimental measurements, rather, the H-GETAWAY descriptors are calculated from the molecular influence matrix H and R-GETAWAY descriptors are derived from the influence/distance matrix R [30]. The diagonal elements (leverages) of the molecular influence matrix H encode atomic information and represent the influence of each atom in determining the whole shape of the molecule. Although the leverage values were shown to be sensitive to significant conformational changes, they account only implicitly for chemical properties of molecule atoms. H-GETAWAY autocorrelation descriptors, e.g., $HATS_k$(w), explicitly allow consideration of chemical information by combining geometrical information provided by leverages with atomic weightings that account for specific physicochemical properties on an atomistic level. The autocorrelation R-GETAWAY descriptors, e.g., maximal R indices R_k^+(w), are analogously defined based on a new matrix R, where the elements of the matrix H are combined with those of the geometry matrix G. R indices are calculated from the autocorrelation at each topological distance (lag) and account for the local aspects of the molecule such as branching, cyclicity and conformational changes. A dependence on conformational changes is expected to increase as *k*-index increases [30,31].

Table 4
Summary of 3D descriptors identified by MCDT as significant with respect to fibrinogen adsorption

Name	Definition	3D class
G(O···O)	Sum of geometrical distances between O···O	Geometrical
G(O···N)	Sum of geometrical distances between O···N	
Mor28(u)	3D-MoRSE signal 28, unweighted	3D-MoRSE
Mor19(e)	3D-MoRSE signal 19, weighted by atomic Sanderson electronegativities	
Mor14(m)	3D-MoRSE signal 14, weighted by atomic masses	
Mor28(m)	3D-MoRSE signal 28, weighted by atomic masses	
Mor3(p)	3D-MoRSE signal 3, weighted by atomic polarizabilities	
Mor7(v)	3D-MoRSE signal 7, weighted by van der Waals volumes	
$HATS_1$(m)	Leverage weighted autocorrelation of lag 1 weighted by atomic masses	H-GETAWAY
$HATS_3$(m)	Leverage weighted autocorrelation of lag 1 weighted by atomic masses	
H_0(e)	H autocorrelation of lag 0 weighted by atomic Sanderson electronegativities	
R_6^+(u)	R maximal autocorrelation of lag 6, unweighted	R-GETAWAY
R_7^+(e)	R maximal autocorrelation of lag 7, weighted by atomic Sanderson electronegativities	
R_5^+(m)	R maximal autocorrelation of lag 5, weighted by atomic masses	
R_3(u)	R autocorrelation of lag 3, unweighted	
R_4(e)	R autocorrelation of lag 4, weighted by atomic Sanderson electronegativities	

Note: leverage is a diagonal element of the molecular influence matrix and lag is defined as a given topological distance.

The fact that specific classes of 3D descriptors were ranked as the best by DT leads one to the conclusion that they encode useful and relevant information with respect to fibrinogen adsorption. It is not a trivial task to relate the descriptors described above with a generalized conformational specificity presented in each of 12 polymers sets. However, one can confidently interpret the appearance in Table 3 of descriptors related to geometrical distances between two oxygen atoms and between oxygen and nitrogen atoms as the importance of electrostatic interactions and hydrogen bonding, respectively. R indices of higher order reflect the conformational changes within each polymer chain while HATS autocorrelations are indicative of the overall shape of the molecule (similar to our case, lower leverages were found in molecules of globular-like or spherical shape [31]). One can also speculate that the presence of 3D-MoRSE signals implicitly confirms the realistic nature of the conformations resulting from MD simulations with emphasis placed on the possible role of electronic structure. In Section 3.3 below the predictive power as well as physicochemical relevance of the descriptors selected by the MCDT will be discussed in the context of the ANN prediction.

3.3. ANN prediction of fibrinogen adsorption

The results of the ANN prediction for the sets of polymers described in Section 3.1 are summarized in Table 5. The accuracy of prediction was evaluated by comparison of the predicted and experimental results for the polymers in the validation set. In contrast to the previous version of ANN where the validation set comprised one half of randomly chosen polymers, the MC-modified ANN employed in the present work allowed for each of 45 polymers to become (at least once) a part of a validation set, and hence, the hypothetical fibrinogen adsorption for all polymers to be estimated. The criteria used for comparison between the prediction and experiment as

Table 5
Results of ANN for the 12 sets of polyarylates introduced in Section 3.1. The error bars represent the standard deviation of the Pearson correlation coefficient due to MC perturbation

3D Set	Vacuum				Water			
	ρ		E_{rms}%		ρ		E_{rms}%	
	Training	Valid.	Training	Valid.	Training	Valid.	Training	Valid.
Minimization								
L	0.80 ± 0.07	0.51 ± 0.22	21	39	0.85 ± 0.06	0.63 ± 0.23	18	38
R	0.79 ± 0.08	0.54 ± 0.19	22	41	0.84 ± 0.07	0.65 ± 0.16	18	35
MD simulations								
L	0.83 ± 0.07	0.66 ± 0.15	18	33	0.83 ± 0.07	0.62 ± 0.13	18	32
R	0.84 ± 0.06	0.63 ± 0.18	18	33	0.85 ± 0.06	0.67 ± 0.13	17	29
Minimization after MD								
L	0.83 ± 0.06	0.59 ± 0.15	18	34	0.84 ± 0.06	0.64 ± 0.15	18	31
R	0.83 ± 0.07	0.61 ± 0.17	20	36	0.84 ± 0.06	0.62 ± 0.17	18	32

well as between the predictions for each polymer set were the percent root-mean-square (rms) relative error E_{rms}, the Pearson coefficient ρ, and the correlation coefficient R^2 (not shown in Table 5) defined by Hawkins [44]. These criteria were prioritized in the same order as they appear above. This prioritization emphasizes the importance for our prediction to be found within or to approach the range of experimental error (i.e., 18%, on average). We also prioritized the Pearson coefficient that ranges from $-1 \leq \rho \leq 1$ in comparison with R^2 that ranges from $-\infty \leq R^2 \leq 1$ and whose lower limit is not clearly defined. Utilization of the Pearson coefficient also allows us a direct comparison with the previously developed surrogate models. We recall that E_{rms} and both correlation coefficients are averaged quantities over 400 MC pseudo-experiments.

From Table 5 one can clearly see that E_{rms} *decreases* in the following order of polymer sets: minimization in vacuum (for L > R isomers) > minimization in water (for R > L isomers) > minimization after MD in vacuum (for L > R isomers) > MD in vacuum (for L = R isomers) > minimization after MD in water (R > L isomers) > MD in water (R > L isomers). The E_{rms} percentage for the last (i.e., the best set with the smallest error) is 29% and compares favorably with experimental error of 18%. This result entirely confirms our hypothesis that MD simulations performed in the presence of water should provide the most realistic conformations of polymers. We concluded that E_{rms} demonstrates the most consistent and reasonable trend.

Surprisingly, the correlation coefficient R^2 reproduces fairly well the tendency shown above. Particularly, R^2 *increases* in the following order: minimization in vacuum (for R > L isomers) > minimization in water (for R > L isomers) > minimization after MD in water (R > L isomers) > minimization after MD in vacuum (for L > R isomers) > MD in vacuum (for R > L isomers) > MD in water (R > L isomers). The Pearson coefficient does not reproduce this order completely; however, the best and the worst polymer sets are the same as those identified using E_{rms} and R^2. For the correlation coefficients the largest inconsistency was noted in the order of performance for two sets of polymers within which the final MD conformations were additionally minimized. This somewhat artificial procedure displayed its important role in evaluation of the sensitivity of 3D descriptors to the small conformational change but possibly did not improve the quality of final conformations by ''forcing'' some relaxed polymer structures into the higher energy local minima. Regarding the sensitivity of 3D descriptors to the conformational diversity of the polymers, one may conclude that the inconsistency of the Pearson coefficient for several polymer sets (as well as the relatively small difference in the magnitudes of E_{rms} and R^2 between sets comprising noticeably different conformations and simulated under distinctive physical settings) is indicative of overall moderate and, for some cases, even low, sensitivity of the 3D descriptors to the conditions investigated in the present study.

Figs. 7a,b and 8a,b,c show selected examples of ANN prediction for validation sets. One can see a significant difference in the accuracy of ANN models (e.g., size of error bars, distribution of polymers along 45° line) for the polymer sets resulting from minimization and MD simulations in vacuum (Fig. 7a and b). A similar trend is also reflected in the magnitudes of standard deviation as well as in the number of outliers that can be noted in Fig. 8a and b for minimization and MD simulations performed in aqueous environment. Fig. 8b and c shows prediction for several chiral polymers numbered accordingly to the convention given in Table 1 as 14, 19, 39, 42 and 43, which were built as L (Fig. 8b) and R (Fig. 8c) isomers with respect to the second chiral center. In this specific example the set containing R isomers has slightly better estimated correlation parameters. Overall chirality-related performance of the polymer sets was not always consistent, though from the trends recorded for E_{rms} and R^2 one can see that overall accuracy of prediction obtained with R isomers is higher than that for L isomers (R > L notation above). Significance of the best and the worst polymer sets in terms of predictive quality sets were successfully identified with all the three criteria employed.

Comparison of the present 3D ANN model with the single 2D ANN model developed by Smith et al. [14] shows noticeable improvement of all estimated parameters. For the best set (i.e., MD simulations in water, R isomers, Fig. 8c) the ANN predicts 78% of polymers in the validation set to be within the range of experimental error in comparison with 70% reported in [14]. The average root-mean-square error for the

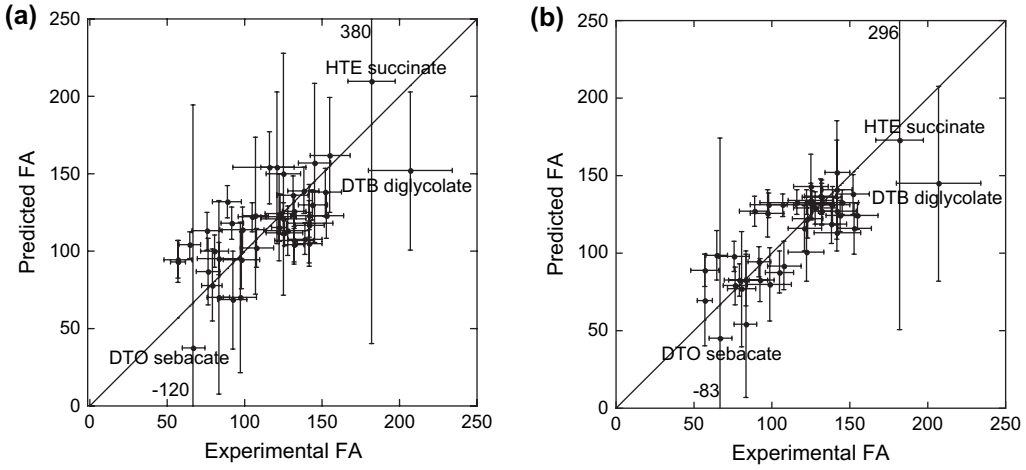

Fig. 7. ANN predictions versus experimental results for validation sets: (a) minimized in vacuum polymer structures; (b) polymer structures from MD simulations in vacuum. In both cases only R isomers are considered. Horizontal error bars represent the standard deviation of the fibrinogen adsorption measurements while vertical error bars represent the standard deviation of the predicted values due to MC perturbation.

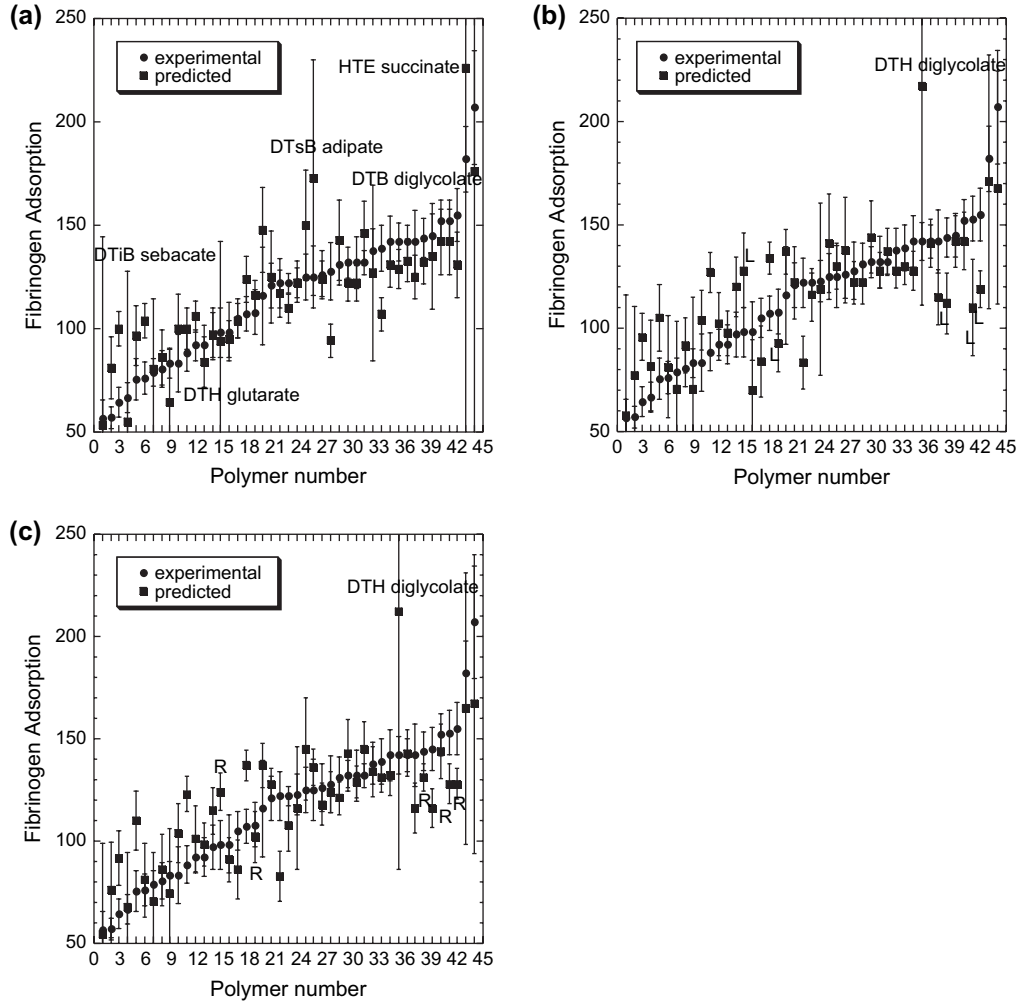

Fig. 8. Validation sets for ANN built for (a) minimized in water polymer structures (only L isomers are included); (b) and (c) conformations from MD simulations in water with L isomers and R isomers included, respectively. Chiral polymers are labeled according to the numbering scheme given in Table 1. Circles represent experimental data and squares correspond to the predicted values. Error bars represent the standard deviation of the fibrinogen adsorption measurements and the standard deviation of the predicted values due to MC perturbation.

validation set prediction decreased from 38% [14] to 29% approaching the value of experimental error (18%), while the average correlation coefficient increased from 0.54 ± 0.12 to 0.67 ± 0.13. Comparison with the prediction obtained for the worst polymer set, namely the structures resulting from the local energy minimization in vacuum, did not show noticeable improvement. In this case, the average root-mean-square error and the average correlation coefficient for the validation set were 41% and 0.54 ± 0.19, respectively. Excellent agreement was observed between the best prediction achieved in the present work and that reported for the extended surrogate model [19]. Both MD-based and extended PLS/PC ANN models exhibit high quality prediction (the average Pearson coefficient of about 0.7) entirely in silico without additional incorporation in the inputs of experimentally measured quantities.

To estimate the quality of our DT analysis we also used a conventional C5 DT routine (without the MC technique) to rank descriptors with respect to fibrinogen adsorption measurements. The results obtained were somewhat surprising: only descriptors from R-GETAWAY group were found to be significant for all 12 sets of polymers. Moreover, the descriptors differ from one another at most by 2 *k*-indices and majority of them were weighted exclusively by van der Waals volume. Further utilization of these descriptors in ANN modeling resulted in very poor prediction (i.e., the Pearson coefficient of 0.2 ± 0.21 and root-mean-square percentage error of 57%) for the polymer set that was shown to be the most successful when MCDT was employed. This fact led us to the conclusion that an application of C5 DT routine does not allow capturing the conformational diversity between the sets of polymers and emphasized the importance of accounting for experimental uncertainty by means of the MC technique.

The ANN results for prediction of fibrinogen adsorption discussed above may pose a question regarding physicochemical relevance on the most significant descriptors listed in Table 3. There are two ways to look at this issue. The conventional way is to consider exclusively descriptors that belong to the first level of significance; in our case, those with the largest counts of MC hits. For instance, appearance of $R_k^+(w)$ indices responsible for the conformational change is associated with the worst ANN model observed for the vacuum minimized polymers structures. $HATS_k(w)$ autocorrelations related to the conformational shape of the molecule, and which in fact belong to the same GETAWAY group as R indices, were found on the first level of DT for the most successful MD-based structural set. The 3D-MoRSE signals weighted by atomic mass and 3D-MoRSE signals weighted by atomic polarizabilities reflect fairly different predictive trends.

One could probably find such interpretation limited and even confusing. Alternatively, when the entire group of the best descriptors is taken into account then more meaningful and consistent information can be extracted. Combination of H-GETAWAY autocorrelations, namely $HATS_1(m)$, $HATS_3(m)$, and geometrical descriptors which encode distances between oxygen atoms, G(O···O), in the most successful polymer sets derived from MD simulations in water indicates the role of electrostatic interactions in polymer folding (see also Table 2) and additionally confirms the realistic nature of this conformational pattern. Appearance of G(O···O) descriptors in conjunction with G(O···N) geometrical distances and 3D-MoRSE signals weighted by atomic polarizabilities indicates a slight shift toward H-bonding and van der Waals interactions [45] in intramolecular energy balance for the polymer conformations resulting from MD simulations in vacuum. It is also not surprising that the intramolecular interactions for this set of polymers are not as crucial as the specific geometrical characteristics which are encoded in a combination of R-GETAWAY and 3D-MoRSE descriptors.

4. Conclusions

A new modeling approach was developed using, for the first time, 3D molecular descriptors obtained for polymer structures relaxed into low energy state by molecular dynamics simulations as inputs into a surrogate model to predict fibrinogen absorption onto polymer surfaces. This 3D surrogate model provided predictions of fibrinogen adsorption that were better than the predictions of the original 2D surrogate model by Smith et al. [14] and comparable to the accuracy of the predictions made by an extended surrogate model recently presented by Smith et al. [19]. The methodology of the extended 2D model described in [19] is mature and has probably been driven to its optimal performance. Further development of this methodology is therefore not expected to lead to substantially better model performance. In contrast, the use of 3D descriptors calculated for realistic structures of the polymers as inputs into surrogate models is new and not yet optimized. This approach holds significant promise and should lead to the development of a model capable of (i) significantly better predictions of key polymer properties, and (ii) prediction of properties of polymers in the absence of experimentally-derived data or descriptors. This last point is of particular importance given the fact that the experimental testing of bioresponses (such as protein adsorption, cell attachment, cell growth, and cell differentiation) will always require laborious experiments that cannot easily be performed for hundreds of polymers.

The present work was performed in three stages. First, MD simulations were carried out for a 45 structurally related polymers of polyarylates library for which experimental fibrinogen adsorption data were previously obtained [21]. Next, 3D descriptors were calculated for molecular structures obtained from (a) preliminary (local) energy minimization, (b) MD simulations in vacuum, and (c) MD simulations in implicit water. Then the Monte Carlo decision tree methodology [14] was utilized to rank the descriptors in order of their correlation to fibrinogen adsorption. Finally, a modified artificial neural network [14,19] was employed for the assessment of 45 polyarylates in terms of their susceptibility to fibrinogen adsorption.

The conformational diversity of the chosen polymers first manifested itself in the minimized structures as a difference in the initial folding of the main chains associated with the presence of two chiral centers in some polyarylates. Despite this, MD simulations indicated that globular-like packing

pattern is typical for all investigated models where the final conformations of the polymers differ noticeably in geometry and configurational energy. Specific intramolecular alignments due to the different flexibilities of diphenol and diacid components of polyarylates are observed. The predictive power and sensitivity of 3D descriptors were estimated using 12 model sets of 45 polymers each. These sets were created to account for the influence of chirality, effect of solvent, and the level of 3D organization (e.g., minimized structures and conformations obtained from MD simulations).

MCDT analysis revealed that the most significant descriptors with respect to fibrinogen adsorption belong to three major groups, namely 3D-MoRSE, GETAWAY and geometrical 3D descriptors. The appearance of these descriptors emphasizes the importance of specific types of intramolecular interactions when polymers fold and form different conformational patterns.

The quality of the final ANN prediction for all polymer sets was estimated using criteria such as the average root-mean-square percentage error and correlation coefficients. These parameters allowed identification of the most successful (MD simulations in water, R isomers) and the least successful (minimization in vacuum L or R isomers) sets of polyarylates. Comparison with the original surrogate model shows that incorporation of descriptors derived from the polymers simulated in implicit water led to a 9% decrease in the average root-mean-square error for the validation set prediction. Simultaneously, the average correlation coefficient increased from 0.54 ± 0.12 to 0.67 ± 0.13. The predictive accuracy of this MD-based model compares favorably with that obtained for the "extended" surrogate model [19]. The prediction obtained for the polymer structures resulting from the local energy minimization is surprisingly close to that of the original ANN model: the average root-mean-square error and the average correlation coefficient for the validation set are 41% and 0.54 ± 0.19, respectively, versus 38% and 0.54 ± 0.12 as reported in Ref. [14].

In summary, we conclude that the 3D surrogate model has four important advantages over the previously described 2D models. This model takes into account more realistic, three-dimensional representation of a polymer; it eliminates experimentally measured descriptors such as air–water contact angle and glass transition temperature; it has better predictive qualities than the original 2D model, and it has significant room for future improvement, based on the utilization of descriptors derived from more sophisticated MD simulations.

Acknowledgments

Financial support for this work was provided by RESBIO, a Research Resource funded by the National Institutes of health under NIH grant EB001046, the New Jersey Center for Biomaterials, and Rutgers University (Research Excellence Award). The authors also acknowledge the help of Ms. Loreto Valenzuela with polymer building and minimization and the help of Dr. Das Bolikal and Dr. Aarti Rege with polymer synthesis.

References

[1] Ratner BD, Hoffman FJ, Schoen JE, Lemons F. Biomaterials science: an introduction to materials in medicine. Elsevier Academic Press; 2004.
[2] Landrum GA, Penzotti JE, Putta S. Meas Sci Technol 2005;16:270–7.
[3] Duce C, Micheli A, Starita A, Tine MR, Solaro R. Macromol Rapid Commun 2006;27(9):711–5.
[4] Bicerano J. Prediction of polymer properties. New York: Marcel Dekker; 2002.
[5] Reynolds CH. J Comb Chem 1999;1:297–306.
[6] Cao C, Lin Y. J Chem Inf Comput Sci 2003;43:643–50.
[7] Yu X, Wang X, Li X, Gao J, Wang H. Macromol Theory Simul 2006;15:94–9.
[8] Tashiro K, Kobayashi M, Yabuki K. Synth Met 1995;71:2101–2.
[9] Linati L, Lusvardi G, Malavasi G, Menabue L, Menziani MC, Mustarelli P, et al. J Phys Chem B 2005;109:4989–98.
[10] Brunswick A, Cavanaugh TJ, Mathur D, Russo AP, Nauman EB. J Appl Polym Sci 1998;68:339–43.
[11] Stamatin L, Berlic C, Vaseashta A. Thin Solid Films 2006;495:312–5.
[12] Ostuni E, Chapman RG, Holmlin RE, Takayama S, Whitesides GM. Langmuir 2001;17(18):5605–20.
[13] Smith JR, Seyda A, Weber N, Knight D, Abramson S, Kohn J. Macromol Rapid Commun 2004;25:127–40.
[14] Smith JR, Knight D, Kohn J, Rasheed K, Weber N, Kholodovych V, et al. J Chem Inf Comput Sci 2004;44:1088–97.
[15] Smith JR, Kholodovych V, Knight D, Welsh WJ, Kohn J. QSAR Comb Sci 2005;24:99–113.
[16] Brocchini S, James K, Tangpasuthadol V, Kohn J. J Am Chem Soc 1997;119(19):4553–4.
[17] Brocchini S, James KS, Tangpasuthadol V, Pendharkar SM, Tong X, Kohn J. Annual meeting of the society for biomaterials. New Orleans, LA; 1997.
[18] Brocchini S, James K, Tangpasuthadol V, Kohn J. J Biomed Mater Res 1998;42:66–75.
[19] Smith JR, Kholodovych V, Knight D, Kohn J, Welsh WJ. Polymer 2005;46:4296–306.
[20] Kholodovych V, Smith JR, Knight D, Abramson S, Kohn J, Welsh WJ. Polymer 2004;45:7367–79.
[21] Weber N, Bolikal D, Bourke SL, Kohn J. J Biomed Mater Res 2004;68(A):496–503.
[22] Schrodinger LLC. Macromodel. v. 8.5. New York, NY; 2005.
[23] Todeschini R, Consonni V, Mauri A, Pavan M. Dragon Web version. v. 3.0. Milano, Italy; 2003.
[24] Research PLR. C5. v. 5.0. St Ives NSW 2075, Australia; 2002.
[25] Still WC, Tempczyk A, Hawley RC, Hendrickson T. J Am Chem Soc 1990;112:6127–30.
[26] Jorgensen WL, Maxwell DS, Tirado-Rives J. J Am Chem Soc 1996;118:11225–36.
[27] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. J Phys Chem B 2001;105(28):6474–87.
[28] Price MLP, Ostrovsky D, Jorgensen WL. J Comput Chem 2001;22(13):1340–52.
[29] Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim (Federal Republic of Germany): WILEY-VCH Verlag GmbH; 2000.
[30] Consonni V, Todeschini R, Pavan M. J Chem Inf Comput Sci 2002;42(3):682–92.
[31] Consonni V, Todeschini R, Pavan M, Gramatica P. J Chem Inf Comput Sci 2002;42(3):693–705.
[32] Murthy SK. Data Min Knowl Discovery 1998;2:345–89.
[33] Shannon CE. Bell System Tech J 1948;27:379–423.
[34] Shannon CE. Bell System Tech J 1948;27:623–56.
[35] Roy NK, Potter WD, Landau DP. IEEE Trans Neural Networks 2006;17(4):1001–14.
[36] Mitchell TM. Machine learning. New York: McGraw-Hill; 1997.

[37] Beale R, Jackson T. Neural computing: an introduction. Bristol, Philadelphia, New York: Adam Hilger IOP Publishing LTD; 1990.
[38] Xiong M, Li W, Zhao J, Jin L, Boerwinkle E. Mol Genet Metab 2001;73:239–47.
[39] Ambroise C, McLachlan GJ. Proc Natl Acad Sci U S A 2002;99(10):6562–6.
[40] Hentschke R. Single chain in solution. In: Kotelyanskii M, Theodorou DN, editors. Simulation methods for polymers. New York, Basel: Marcel Dekker Inc; 2004.
[41] Nakamura K, Houk KN. Org Lett 1999;1(13):2049–51.
[42] Gasteiger J, Sadowski J, Schuur J, Selzer P, Steinhauer L, Steinhauer V. J Chem Inf Comput Sci 1996;36(5):1030–7.
[43] Schuur JH, Selzer P, Gasteiger J. J Chem Inf Comput Sci 1996;36(2):334–44.
[44] Hawkins DM. J Chem Inf Comput Sci 2004;44:1–12.
[45] Carrasco R, Padron JA, Galvez J. J Pharm Pharm Sci 2004;7(1):19–26.