Polymer 49 (2008) 2435–2439

Contents lists available at ScienceDirect

Polymer

journal homepage: www.elsevier.com/locate/polymer

Polymer Communication

Prediction of biological response for large combinatorial libraries of biodegradable polymers: Polymethacrylates as a test case

Vladyslav Kholodovych [a,*,1], Anna V. Gubskaya [b,1], Michael Bohrer [b], Nicole Harris [b], Doyle Knight [c], Joachim Kohn [b], William J. Welsh [a,*]

[a] Department of Pharmacology, Robert Wood Johnson Medical School (RWJMS), University of Medicine and Dentistry of New Jersey (UMDNJ), Piscataway, NJ 08854, United States
[b] New Jersey Center for Biomaterials, Rutgers, The State University of New Jersey, New Brunswick, NJ 08854-8087, United States
[c] Department of Mechanical and Aerospace Engineering, Rutgers, The State University of New Jersey, New Brunswick, NJ 08854-8058, United States

ARTICLE INFO

Article history:
Received 29 November 2007
Received in revised form 11 March 2008
Accepted 19 March 2008
Available online 27 March 2008

Keywords:
Combinatorial chemistry
Biomaterials
Computer modeling

ABSTRACT

A large virtual combinatorial library of polymethacrylates was, for the first time, designed for computer-aided prediction of biorelevant and material properties and focused polymer synthesis. The distinguishing features of this virtual library include its size (about 40 000 compounds), its explicit representation of relatively long polymer chains, and its accounting for different compositions in the case of copolymers and terpolymers. A subset of 79 polymers taken from a representative sub-library of 2000 polymethacrylates was employed to build initial QSPR-based polynomial neural network models, which were then deployed to predict cell attachment, cell growth, and fibrinogen adsorption on polymer surfaces for these 2000 polymethacrylates. The agreement between predicted and experimentally measured property values for the 50 polymethacrylate copolymers within this virtual polymer space encourages further pursuit of polymethacrylate-based biomaterials, and justifies more extensive deployment of computational models derived from larger experimental data sets for the rational design of biorelevant polymers endowed with targeted performance properties.

1. Introduction

Utilization of computational molecular modeling methodologies has led to significant accomplishments in a wide variety of fields of research and development. Computational approaches and tools employed successfully in pharmaceutical drug discovery, such as molecular diversity/similarity, virtual screening, and quantitative structure–performance relationship (QSPR) models, have become essential technologies over the past decade due to their scalability, robustness and predictability. Recent advances in QSPR models derived from machine-learning algorithms, together with the advent of affordable high-performance computer hardware and software, now invites extension of these methodologies to larger and more complex systems such as biomaterials. Nevertheless, computational modeling and prediction of bioresponse phenomena for polymeric biomaterials remains a significant challenge.

The first combinatorial library of polymers [1] consisted of 112 strictly alternating polyarylate copolymers, and beginning in 1999 [2], has served as the source of data for building an array of predictive QSPR (surrogate or semi-empirical) models that differed with respect to their theoretical basis and computational complexity. Statistically robust QSPR models were constructed and validated to predict diverse physicochemical and biological properties such as glass transition temperature (T_g), air–water contact angle (AWCA), fibrinogen adsorption onto polymer surfaces [3,4], and cellular attachment and growth of fibroblast-like cells from mouse embryo on flat surfaces of the selected polyarylates [5]. Several computational techniques, such as multiple linear regression (MLR), partial least square analysis (PLS) [5] and artificial neural networks (ANN) [3,6], have been employed and compared with respect to their statistical quality and predictability. Despite some qualitative differences in their predictability, all of these models were reasonably successful in predicting the properties of the "external" set of polymers, i.e., the polymers, which were not involved in generating the QSPR models. These previous computational efforts employed various strategies to select molecular structure-based descriptors. Among these approaches, the decision tree algorithm, the Monte Carlo variation procedure, and principal component analysis (PCA) [4,6] were successful in selecting molecular descriptors that correlated most strongly with experimentally determined performance properties. While early models included experimental values of T_g and AWCA as descriptors for

* Corresponding authors. Fax: +1 732 325 3475.
E-mail addresses: kholodvl@umdnj.edu (V. Kholodovych), welshwj@umdnj.edu (W.J. Welsh).
[1] These authors contributed equally to the present work.

0032-3861/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2008.03.032

building QSPR models [3], more recent work has shown that highly predictive QSPR models could be derived solely from computationally generated descriptors [4,5]. Physical interpretation of the most significant descriptors enabled these researches to establish structure–performance relationships among polyarylates and to formulate principles for guiding optimal polymer design. The success of these past efforts confirmed the advantages in integrating experimental and computational methods toward the rational design of polymers for specific applications.

Where QSPR models have been constructed, the ANN algorithm has been the tool of choice in the majority of cases [5–9]. In contrast, the present study employs a distinct type of neural network called the polynomial neural network (PNN) [10]. Despite their similar composition in terms of multi-layered interconnectivity of neurons (nodes), the ANN and PNN differ fundamentally with respect to their dynamic nature. The ANN requires the user to specify the architecture, i.e. the precise number of neurons in the input, output and hidden layers, as well as a number of hidden layers. In contrast PNN assembles the architecture in response to features of the data [11]. It can generate both linear and non-linear regression equations of any order and automatically produces any number of QSPR equations that best correlate the inputs (molecular descriptors) and outputs (target properties) and sorts them in order of their predictive ability. Furthermore, the PNN is specifically designed to process both very small data sets and very large (>5000 compounds) data sets, even when the data are "noisy" or contain irrelevant values (outliers). These features make PNN an invaluable tool for QSPR studies of large and diverse combinatorial libraries of polymers.

The goal of this communication is threefold: (a) to introduce the major advantages of automated parallel synthesis of polymethacrylates using the reversible addition–fragmentation chain transfer polymerization technique; (b) to present the design of a large virtual combinatorial library of polymethacrylates where each compound is represented by an explicit chain of repeat units assembled to reproduce several polymeric compositions; and (c) to report initial PNN models for the polymethacrylates that show promise for predicting of the bioresponse of large and diverse combinatorial libraries of polymers. The present study illustrates a rational strategy for screening and selection of the most promising candidates for focused parallel synthesis.

2. Methods

2.1. Automated synthesis

From the perspective of combinatorial polymer synthesis polymethacrylates present an attractive choice due to the broad range of structural diversity in the pendant ester group of commercially available monomers (see Table S1 in Supplemental materials). Additionally, these monomers are predominantly liquids, which make them amenable to the superb liquid handling of the automated synthesizer [12]. The polymers were synthesized as homopolymers, copolymers, and terpolymers using reversible addition–fragmentation chain transfer (RAFT) polymerization in order to obtain well-defined polymers (Fig. 1). RAFT polymerization is a robust technique, which is amenable to many monomers and solvents [13]. Since the control of the polymerization relied heavily on the RAFT chain transfer agent, 2-cyanoprop-2-yl dithiobenzoate was used to this end and it exhibited good control for the synthesis of various methacrylate monomers [12,14]. In a second step, the RAFT end group was exchanged for an isobutyronitrile end group by reaction with an excess of azobisisobutyronitrile (AIBN) [15,16]. Automated synthesis allowed for rapid optimization of reaction conditions, including solvent, concentration, and ratio of RAFT chain transfer reagent to free radical initiator (Fig. 1). In total 79 polymers from the combinatorial library of polymethacrylates were synthesized; their thermal and biological properties were determined and these polymers comprised the training set for subsequent computational modeling. To our knowledge, the synthesized subset of polymethacrylates is the first polymer library obtained using automated synthesis with the aim to generate the structure–activity relationships necessary to tune the properties of biomaterials.

2.2. Biological characterization

The advent of combinatorial parallel synthesis of biomaterials prompted the need to develop efficient rapid-screening assays to investigate the diversity of responses when biological systems (i.e., macromolecules, cells, organs, whole organisms) are exposed to the polymeric material. A rapid-screening immuno-fluorescence assay (IFA) for the detection of fibrinogen adsorption onto polymer surfaces developed by Weber et al. [17] was validated on polymethacrylates and employed in the present study. An established procedure [18] was utilized to culture NIH3T3 (fibroblast-like cells from mouse embryo) on flat surfaces of the selected polymethacrylates. Metabolic activity represented by cell attachment and cell growth was estimated using a commercially available MTS assay (CellTiter96®, Promega, Madison, WI). Altogether three different types of bioresponses, specifically, fibrinogen adsorption (FA), cell attachment (CA), and cell growth (CG) on the polymer surfaces were measured for the selected subset of polymers. The resulting set of experimental data, together with the calculated molecular descriptors, was employed to build QSPR models using PNN.

2.3. Polynomial neural network

The PNN is a powerful machine-learning algorithm [10] that generates both linear and non-linear QSPR regression models in parametric form. By virtue of its ability to handle either sparse or large datasets, the PNN was chosen to build predictive models for this virtual combinatorial library of polymethacrylates. This unique feature of the PNN was deemed especially useful in the present study, which sought to predict the bioresponse for a large (up to 2000) and diverse data set of polymethacrylates based on QSPR models built using experimental measurements for the data set that consisted of 79 polymers. Only polymers whose experimentally measured standard deviation was within 25% of their mean value were selected for model construction. To avoid the problem of

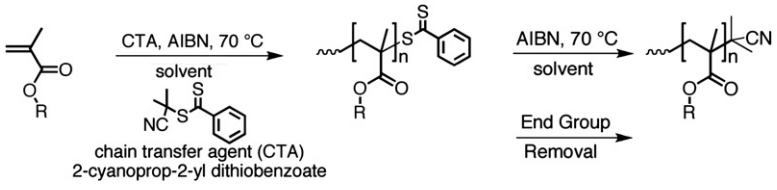

Fig. 1. Polymethacrylate library: reversible addition–fragmentation transfer (RAFT) polymerization.