

## Polymer Communication

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

Vladyslav Kholodovych<sup>a,\*</sup>, Anna V. Gubskaya<sup>b,1</sup>, Michael Bohrer<sup>b</sup>, Nicole Harris<sup>b</sup>, Doyle Knight<sup>c</sup>, Joachim Kohn<sup>b</sup>, William J. Welsh<sup>a,\*</sup>

<sup>a</sup> Department of Pharmacology, Robert Wood Johnson Medical School (RWJMS), University of Medicine and Dentistry of New Jersey (UMDNJ), Piscataway, NJ 08854, United States

<sup>b</sup> New Jersey Center for Biomaterials, Rutgers, The State University of New Jersey, New Brunswick, NJ 08854-8087, United States

<sup>c</sup> 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.

© 2008 Elsevier Ltd. All rights reserved.

### 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](mailto:kholodvl@umdnj.edu) (V. Kholodovych), [welshwj@umdnj.edu](mailto:welshwj@umdnj.edu) (W.J. Welsh).

<sup>1</sup> These authors contributed equally to the present work.

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<sup>®</sup>, 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

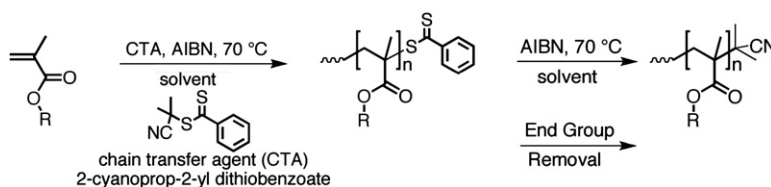


Fig. 1. Polymethacrylate library: reversible addition–fragmentation transfer (RAFT) polymerization.

model over-fitting, each PNN-based QSPR model was restricted to a maximum of five descriptors from the pool of 184 descriptors.

### 3. Results and discussion

#### 3.1. Design of virtual combinatorial library

To introduce structural diversity into our combinatorial design, 33 building blocks (side chains) from a family of methacrylates were chosen as candidates for “computational combinatorial synthesis”. A length of the common structural skeleton (polymer backbone) was built to accommodate 12 repeat units to encompass variations in both copolymers and terpolymer compositions. The molecular modeling environment (MOE) molecular modeling package [19] was used to generate a virtual library of about 40 000 polymers, where the building blocks were randomly combined to reproduce 50/50, 25/75, 75/25, and 33/33/33 percent compositions for copolymers and terpolymers. The entire combinatorial virtual library included all possible permutations and comprised three sub-libraries: (a) homopolymers of 33 polymers; (b) copolymers of 3267 ( $33 \times 33$ ) polymers and (c) terpolymers of 35 937 ( $33 \times 33 \times 33$ ) structures. Schematic representation of this virtual library is shown in Fig. 2. The binomial coefficient [20] was used to select unique polymers by reducing the total number of copolymer structures from 1089 to 528 for each of the 50/50, 25/75 and 75/25 percent compositions (total 1584 compounds). Similarly, the 35 937 terpolymers were reduced to 5456 unique structures and the data set of 330 terpolymers was employed in modeling part of this study (see Fig. 2). Exceptionally large size, structural diversity, and explicit representation of polymer compositions make this virtual library of polymethacrylates an ideal test case for computer-aided prediction of biorelevant and material properties.

The MMFF force-field [21] was used to generate energy-minimized structures of each polymer under consideration. The MMFF contained a complete set of force-field parameters for all the atoms present in monomers of polymethacrylates chosen to build the library (Table S1). Two-dimensional descriptors that encode for molecular connectivity (184 descriptors altogether for each polymer) were computed using the MOE software. It has been previously shown that these descriptors are capable of successfully predicting similar types of bioresponse for the library of biodegradable polyarylates [3,5].

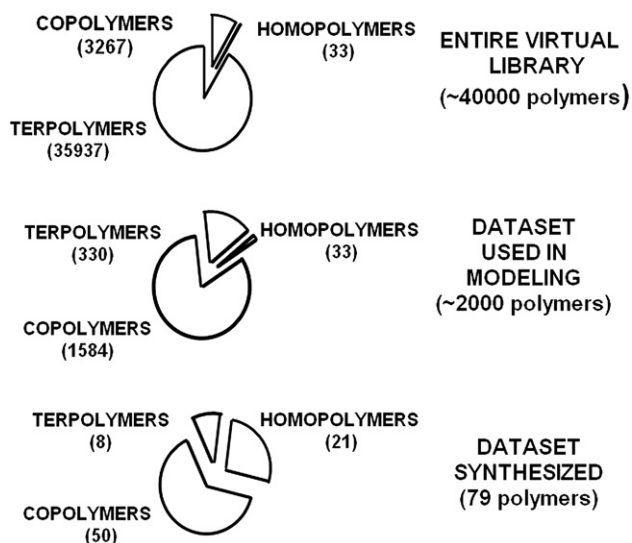


Fig. 2. Schematic representation of the virtual library of polymethacrylates.

Table 1

Values of the correlation coefficient ( $R^2$ ) and mean absolute error (MAE) from the PNN models generated for all copolymers from the virtual combinatorial library of polymethacrylates

	Cell attachment	Cell growth	Fibrinogen adsorption
Total number of polymers	22	20	22
Split of training and test sets	17/5	16/4	18/4
$R^2$ (training set) model I	0.80	0.95	0.86
$R^2$ (combined) model II	0.78	0.95	0.78
MAE (training set) model I	14.67	8.01	10.50
MAE (combined) model II	17.90	10.94	15.46

#### 3.2. PNN models

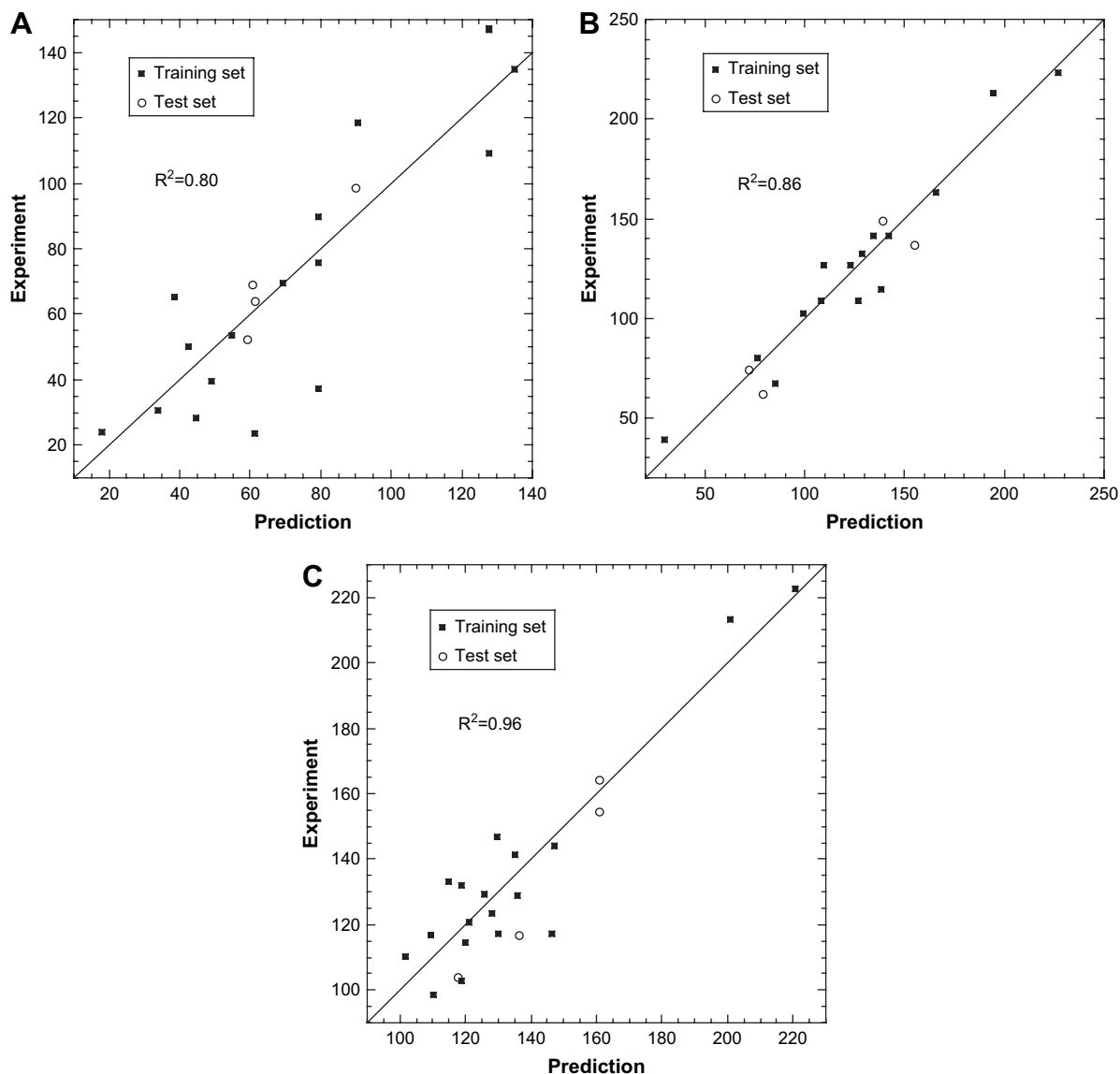
To build PNN models for copolymers the three experimental data sets generated for separate compositions (i.e., 50/50, 75/25, 25/75) were combined into a single data set, which provided a sufficient number of data points to build a statistically robust model and, at the same time, to cull a representative test set of copolymers for model validation (Table 1, Fig. 3).

Two linear regression models, specifically, model I based solely on the training-set compounds and model II based on the combination of training and test sets, were generated (Table 1). Model II was subsequently employed to predict the bioresponse for each polymer in the entire virtual library of ~2000 polymethacrylates. Values of the correlation coefficient ( $R^2$ ) for all training-set models were of the order of 0.80 or higher. Agreement between the model predicted and experimentally measured values was excellent for the test set polymers, aside from the cell attachment data. However, exclusion of just one compound from the test set substantially improved the predictive ability of this model (Table 2). Further prediction was made for the selected combinatorial library of polymethacrylates comprising 33 homopolymers, 1584 copolymers at 50/50, 25/75 and 75/25 percent compositions and a representative subset of 330 terpolymers at 33/33/33 percent. Those polymers with favorable bioresponse values were selected as potential candidates for subsequent chemical synthesis and biological evaluation. In view of the positive results obtained for the copolymers, future work will expand the paradigm to the entire library of terpolymers.

Analysis of the molecular descriptors selected by PNN revealed that hydrophobic/hydrophilic features of the polymers correlated strongly with their biological behavior. Descriptors related to the octanol/water partition coefficient ( $\log P$ ) contributed significantly to the QSPR model for all three types of bioresponses. Two additional descriptors selected by PNN for cell attachment where electrostatic interactions are crucial for the attraction of cells were the number of nitrogen atoms and the net charge of the polymer. The salient descriptors for the cell growth model, in addition to the contribution from  $\log P$ -related descriptors, included a shape coefficient and selected electron donor groups; the latter probably plays a role in locking growing cells to the polymer surface over time. Among the top five descriptors selected by PNN for the fibrinogen adsorption model was the number of rotatable bonds, suggesting that the conformational flexibility of the polymers is vital for interactions with protein.

### 4. Conclusions

We report here the computational combinatorial design of a large virtual library of polymethacrylates together with initial PNN models, which were built to predict clinically relevant bioresponses for diverse polymer designs including copolymers and



**Fig. 3.** Model II-type PNN models for a set of 22 polymethacrylates: (A) cell attachment; (B) cell growth; (C) fibrinogen adsorption. Correlation coefficients are shown for training sets.

terpolymers. The results obtained offer guidance for the rational design of polymethacrylate-based biomaterials, and promote further development and application of computational models based on more extensive experimental data sets for this family of polymethacrylates. Although the number of experimental data points for model generation was rather sparse, the resulting models

demonstrated good quality correlations and allowed quantitative performance prediction for the much larger set of about 2000 representative polymethacrylates. The polymers identified by combined regression models as “active” have been selected for synthesis and biological characterization to enable further validation of the initial PNN models.

**Table 2**  
Validation results for the linear regression PNN model

Cell attachment <sup>a</sup> (% of control)		Cell growth <sup>a</sup> (% of control)		Fibrinogen adsorption <sup>b</sup> (% of control)	
Experiment	Prediction	Experiment	Prediction	Experiment	Prediction
<b>93.7</b>	<b>54.8</b>	137.0	154.5	164.3	160.8
98.8	89.9	74.2	71.9	104.1	117.50
52.5	59.0	149.2	138.8	154.5	160.9
64.1	61.1	62.1	78.8	116.7	136.4
69.2	60.4	–	–	–	–

Experimentally measured and predicted values of cell attachment, cell growth and fibrinogen adsorption are shown for test sets of 5, 4 and 4 polymethacrylates, respectively; compound depicted in bold was excluded from the test set.

<sup>a</sup> Relative to tissue culture polystyrene (TCPS) [5].

<sup>b</sup> Relative to polypropylene control [17].

## 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 assistance of Dr. Karolina Piotrowska for the polymer synthesis and Drs. Paul Holmes and Prafulla Chandra for help with the biological characterization.

## Appendix. Supplementary data

List of polymethacrylate monomers used to build virtual combinatorial library is provided in Supplementary data.

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.polymer.2008.03.032.

## References

- [1] Brocchini S, James K, Tangpasuthadol V, Kohn J. *Journal of Biomedical Materials Research* 1998;42:66–75.
- [2] Reynolds CH. *Journal of Combinatorial Chemistry* 1999;1:297–306.
- [3] Smith JR, Knight D, Kohn J, Rasheed K, Weber N, Kholodovych V, et al. *Journal of Chemical Information and Computer Science* 2004;44:1088–97.
- [4] Smith JR, Kholodovych V, Knight D, Kohn J, Welsh WJ. *Polymer* 2005;46:4296–306.
- [5] Kholodovych V, Smith JR, Knight D, Abramson S, Kohn J, Welsh WJ. *Polymer* 2004;45:7367–79.
- [6] Smith JR, Kholodovych V, Knight D, Welsh WJ, Kohn J. *QSAR and Combinatorial Science* 2005;24:99–113.
- [7] Duce C, Micheli A, Starita A, Tine MR, Solaro R. *Macromolecular Rapid Communications* 2006;27(9):711–5.
- [8] Roy NK, Potter WD, Landau DP. *IEEE Transactions on Neural Networks* 2006;17(4):1001–14.
- [9] Selzer P, Ertl P. *Journal of Chemical Information and Modeling* 2006;46(6):2319–23.
- [10] Aksenova TI, Volkovich V, Tetko I. *SAMS* 2003;43:1331–41.
- [11] Ivahnenko AG. *IEEE Transactions on Systems, Man and Cybernetics* 1971;SMC-1(4):364–78.
- [12] Fijten MWM, Paulus RM, Schubert US. *Journal of Polymer Science, Part A: Polymer Chemistry* 2005;43:3831–9.
- [13] Chiefari J, Chong YK, Ercole F, Krstina J, Jeffery J, Le TPT, et al. *Macromolecules* 1998;31:5559–62.
- [14] Moad G, Rizzardo E, Thang SH. *Australian Journal of Chemistry* 2005;58:379–410.
- [15] Perrier S, Takolpuckdee P, Mars CA. *Macromolecules* 2005;38:2033–6.
- [16] Yanjarappa MJ, Gujraty KV, Joshi A, Saraph A, Kane RS. *Biomacromolecules* 2006;7:1665–70.
- [17] Weber N, Bolikal D, Bourke SL, Kohn J. *Journal of Biomedical Materials Research* 2004;68(A):496–503.
- [18] Brocchini S, James K, Tangpasuthadol V, Kohn J. *Journal of the American Chemical Society* 1997;119(19):4553–4.
- [19] Chemical Computing Group Inc. MOE (The molecular operating environment). Montreal, Canada H3A 2R7; 2005.
- [20] Rosen KH. *Handbook of discrete and combinatorial mathematics*. CRC Press LLC; 2000.
- [21] Halgren TA. *Journal of Computational Chemistry* 1999;20(7):730–48.