Published in final edited form as:

Chem Biol Drug Des. 2012 April ; 79(4): 478–487. doi:10.1111/j.1747-0285.2011.01293.x.

Virtual Design of Chemical Penetration Enhancers for Transdermal Drug Delivery

Sharath Golla, Brian J. Neely^{*}, Eric Whitebay, Sundar Madihally, Robert L. Robinson Jr, and Khaled A. M. Gasem

All authors at School of Chemical Engineering, Oklahoma State University, Stillwater, OK

Abstract

Traditional drug design is a laborious and expensive process that often challenges the pharmaceutical industries. As a result, researchers have turned to computational methods for computer-assisted molecular design. Recently, genetic and evolutionary algorithms have emerged as efficient methods in solving combinatorial problems associated with computer-aided molecular design. Further, combining genetic algorithms (GAs) with quantitative structure-property relationship (QSPR) analyses has proved effective in drug design.

In this work, we have integrated a new genetic algorithm and non-linear QSPR models to develop a reliable virtual screening algorithm for generation of potential chemical penetration enhancers (CPEs). The GA-QSPR algorithm has been implemented successfully to identify potential CPEs for transdermal drug delivery of insulin. Validation of the newly-identified CPE molecular structures was conducted through carefully designed experiments, which elucidated the cytotoxicity and permeability of the CPEs.

Keywords

Virtual Screening; Structure-Based Drug Design; Drug Design

1. Introduction

The demand for newly-designed molecules that enhance existing processes and satisfy more stringent operating requirements in technology has been increasing. However, the rational design of molecules with desired properties challenges engineers attempting to meet the needs of various industries, including pharmaceuticals, polymers, petrochemicals, and construction (1-3). The traditional approach of identifying molecules with desired properties involves testing thousands of molecules for their chemical and physical properties, which is an expensive and laborious undertaking. Hence, rational design techniques, such as computer-aided molecular design (CAMD), have found wide application in recent years (4, 5).

CAMD methods have been employed successfully in a wide range of applications, including solvent design/selection (6), chloro-fluoro-carbon (CFC) substitutes, alternative process fluids, polymer design (2), drug design (7), and design for novel molecules with superior properties (4). A typical CAMD design algorithm utilizes two key components, (a) a method for generating candidate molecules and (b) models to predict the pertinent physiochemical properties of the newly generated molecules. Property predictions for the generated molecules are usually done using group contribution methods, equation-of-state approaches,

^{*}corresponding author: neelyb@okstate.edu, 405-744-5280.

and quantitative structure-property relationship (QSPR) models. Figure 1 presents a general view of the various stages, which are involved in CAMD.

In this work, a combination of genetic algorithms and QSPR techniques has been used to develop the CAMD algorithm for virtual design of chemical penetration enhancers (CPEs) for transdermal drug delivery. This work is in response to the extensive efforts that have been expended in search of chemicals to enhance the penetration of therapeutic drugs through human skin. Although CPEs can be valuable in increasing the amount and/or rate of drug delivery, they can also have undesirable effects, including skin irritation and toxicity. Thus, a distinct need exists for effective methods to identify new CPEs that provide optimum penetration enhancement with minimal side effects.

The primary goal of the work reported here was the integration of non-linear QSPR modeling and robust genetic algorithms (GAs) to facilitate the rational design of improved CPEs. Our basic premise is that novel, effective mathematical models can be developed to describe accurately the relationship between the molecular structure of a chemical and its CPE behavior, and that these models can form the basis for the “virtual design” of promising molecular structures for use as CPEs. Ultimate benefits of such a design capability include (a) identifying novel CPEs, (b) reducing the need for expensive and time-consuming experiments, and (c) setting the stage for the synthesis and commercialization of improved CPEs for use by the medical community.

The work proceeded in distinct stages, as described below. To begin, the target properties of CPEs were identified by a thorough literature survey and analysis of their molecular properties. Using a database of 272 CPEs cited in the literature as seed molecules, new molecules were generated using genetic operators such as crossover, mutation and functional group addition. QSPR models developed using artificial neural networks (ANNs) were used to predict the target physiochemical properties, including skin penetration coefficient (8), octanol/water partition coefficient, melting point (9), skin sensitization and skin irritation (10, 11) of the newly generated molecules. The molecules were scored and screened before being passed to the next generation. To further validate the design methodology results, all identified potential CPEs were tested experimentally for toxicity and skin permeation through carefully designed measurement techniques, which are described in detail elsewhere (12-15).

2. Computer-Aided Molecular Design Methods

In traditional sequential methods of molecular discovery for developing new chemicals, often several hundred (and in the case of drug design, several thousand) new molecules may be tested experimentally before a viable chemical is identified. An attractive alternative to such development schemes is the use of virtual screening, wherein (a) the physical synthesis of molecules is replaced by virtual synthesis, (b) the experimental property measurements are minimized through the use of accurate property prediction models, and (c) robust scoring modules guide the virtual screening algorithms toward the most feasible subset of molecules. The complexity of CAMD calculations depend upon the targeted application and the computational techniques used.

CAMD techniques have been implemented successfully by our research group for more than a decade, focusing on the design of solvents for extractive distillation (6, 16). We recognized that our third-generation chemical design algorithms developed for design of proprietary solvents (17) could be applied effectively in CPE design, once calibrated properly for this application. A CAMD problem typically involves the following steps, as proposed by Gani and coworkers (4, 18-20) (these steps are described in greater detail in the following sections):

- Problem formulation – The target physiochemical properties and their desired values (or ranges of values) are determined.
- Initial search – Molecules identified as potential CPEs in the literature are introduced into the CAMD algorithm as seed molecules.
- Molecule generation and testing – Using the seed molecules, new molecules are generated and tested.
- Verification – The efficacy of the selected molecules is validated experimentally.

2.1 Problem formulation

Identifying the desired (target) properties of the chemical compounds to be generated is the defining step in CAMD processes. In this work, the target molecules should enhance the permeation of a selected drug through the skin without causing any harmful effects. After thorough analysis of the currently available CPEs and their properties, the following properties were identified as being significant for transdermal drug delivery:

1. Molecular weight: Molecules with low molecular weights easily penetrate the skin due to their small size. Hence an upper limit of 500 was imposed on the molecular weight of potential CPEs (21-24).
2. Octanol/water partition coefficient (K_{ow}): Drugs with very low or high partition coefficient fail to reach systemic circulation (22-24). Several ranges of log K_{ow} values have been proposed in the literature for effective permeation enhancement. In this work, molecules with log K_{ow} values in the range of 1-3 were accepted and considered to indicate good permeation enhancement (21).
3. Melting point: Molecules with high melting points, due to their low solubility both in water and fat, are ineffective in transdermal drug delivery (TDD) (22), and only molecules with melting points less than 200° C were accepted (21).
4. Skin sensitization: The CPE should not cause any skin irritation or sensitization upon application (21). All the newly generated molecules are scored using three independent skin sensitization QSPR models, and only those molecules that are classified as being non-sensitizers by all three models were passed to the next generation.
5. Number of hydrogen donor groups: The sum of the hydrogen atoms linked to oxygen and nitrogen atoms in the molecule determines the total number of hydrogen-bond donor groups in a molecule. The permeability across the lipid bi-layer has been identified to be significantly lower for drugs with an excessive number of these groups (21, 24). Hence, a hydrogen-bond donor number upper limit of five was specified for acceptance of a molecule as a CPE.
6. Number of hydrogen-acceptor groups: The total number of nitrogen, oxygen and fluorine atoms in the molecule (excluding nitrogen atoms with a formal positive charge, higher oxidation states and pyrrolyl forms) determines the total number of hydrogen-bond acceptor groups in a molecule. Presence of too many acceptor groups has been identified as a hindrance to the permeability across the lipid bi-layer (24), and therefore an upper limit of 10 was used for the hydrogen-bond acceptor number.

2.2 Initial search

The chemical structures (genetic material) already known to be effective CPEs were utilized by the GA to generate new potential chemical structures. Accordingly, a thorough literature

search was required to assemble available CPE data. A literature search focused on database compilation of CPE molecules was completed by Osborne and Henke (25), where over 400 technical and patent literature sources were reviewed, and a dataset of 275 CPE molecules was compiled. Only molecules that enhance skin permeability by reversibly altering the skin or by changing the physiochemical nature of the skin were included in the database.

The chemical classes present in the database include fatty alcohols, fatty acids, fatty acid esters, fatty alcohol ethers, biologics, enzymes, amines, amides, complexing agents, macrocyclics, classical surfactants, pyrrolidones, ionic compounds, solvents and azone-related compounds. In a recent article involving over 90 technical and patent literature sources, Thong *et al.* (26) studied CPE classification and mechanisms and provided a database of approximately 180 CPE molecules, along with their chemical classes, mechanisms of action and examples of targeted drugs.

These two databases were studied carefully and a new database (Oklahoma State University [OSU] CPE database) consisting of over 400 CPE molecules from diverse chemical classes was compiled. Of these 400 CPEs, molecular structures of 272 CPEs were identified using multiple software applications, and used as seed molecules in our GA approach for CPE design.

2.3 Molecular generation and testing

Genetic algorithms—While the desired properties and their target values, as well as the list of candidate molecules, depend on the specific CAMD application, the efficiency of the CAMD technique depends on the methods used for molecule generation and property evaluation (5).

GAs function by providing new molecules in each generation through crossover and mutation operations applied to randomly-selected candidate molecules. All newly generated molecules undergo a scoring process where molecules are assigned a numerical score based on their property values. These molecules are screened, and those scoring well are passed to the next generation. Figure 2 summarizes the GA methodology for CPE design.

Molecular representation—Developing a GA for CPE molecular design requires an effective molecular representation scheme. Various methods for molecular representation are used in practice (27-29). The Simplified Molecular Input Line Entry System (SMILES) notation used in this work is a line/string notation that is human readable and can be transformed easily into a 2-D structure (27). In this work, all seed molecules were converted to SMILES notation using OpenBabel software (30).

Genetic operators—GAs involve random selection of parent molecules to generate new offspring. To accomplish this, a variety of genetic operators and processes are used, as discussed below.

1. **Selection:** The genetic algorithm has been designed on the basis of a probabilistic operator, rather than a deterministic one used by other optimization techniques. This means that the molecular growth is completed in a random fashion with a greater probability of selection given to those molecules possessing superior characteristics, as determined by the fitness function. This is achieved by using what is called “Roulette Wheel Selection” (31).
2. **Crossover:** The crossover operator creates an offspring by combining features of the parent molecules. Figure 3 shows two crossover operators: one-point crossover and two-point crossover. Roulette wheel selection is used to select between one-point and two-point crossovers in each generation.

3. Mutation: The mutation operator performs random changes in the parent molecule to produce a new offspring. Figure 4 presents an example of the various mutation operators used.
4. Other operators: Other genetic operators used for molecular generation include functional group addition, functional group deletion and bond substitution. Figure 4 includes examples of these operators. The functional groups to be added are selected from a pool of functional groups identified as being prominent in currently available CPEs.

Development of fitness function—Scoring and screening of generated molecules is a key step in any CAMD technique. A GA-based search technique typically analyzes a few thousand molecules before a suitable candidate is identified. Several techniques have been developed for scoring and screening of generated molecules. One such method is the rejection of candidates that do not satisfy the target property constraints. This method is effective only when the feasible region in the search space is large. All generated molecules are given a fitness score using a fitness function that is tailored to a specific problem. The fitness score can be evaluated in two ways:

1. Assign a score to the molecule based on predicted property values.
2. Specify an acceptable range for each of the properties under consideration.

Each of these methods has advantages. By giving a score to each of the molecules through a set of property models, a minimum score for acceptance can be specified; thus, molecules are not rejected for violating one or more of the property constraints. We believe a combination of these two approaches -- scoring of the initial generations and specifying a range for the final generations -- provides an effective fitness evaluation routine, and this method was used in the present study.

2.4 Verification

Careful experimental validations for both skin permeation and toxicity were conducted on those identified candidate CPEs that demonstrated the greatest potential. Details on the experimental validation capability of the OSU Chemical Design Research Group are beyond the scope of this study and are given elsewhere (12-15).

3. Results and Discussion

3.1 QSPR models

QSPR models for properties, including skin penetration coefficient ($\log K_p$) (8), octanol/water partition coefficient ($\log K_{ow}$), melting point (9), and skin sensitization (10, 11), were developed to predict the physiochemical properties of the newly-generated molecules. The steps in QSPR modeling include assembly and characterization of a suitable database, characterization of the molecular structures through molecular descriptor calculation and finally, development of a QSPR model.

To ensure that the QSPR models have reliable prediction capabilities, molecular databases consisting of chemicals from diverse chemical classes and spanning wide property ranges were used for model development. Molecular descriptors, which are calculated based on the molecular structure, are used to characterize the chemicals comprising the assembled databases. An important aspect in calculation of these descriptors is accurate representation of the molecular structure. The chemical structures used for modeling were optimized initially using the Chem3D module available in Chem3DUltra (32). To increase confidence that the lowest energy configuration was located, multiple initializations were used during

the structure optimization. AMPAC 6.0 software (33) was then used to further refine the 3-D geometry of the structures. The final optimized structures from AMPAC were provided as inputs to commercial QSPR software to calculate over 1200 molecular descriptors. A variety of constitutional, topological, geometrical, thermodynamic, quantum-chemical and electrostatic descriptors were calculated using CODESSA (34) and 154 functional group descriptors were calculated using Dragon (35). The number of descriptors calculated for each molecule depended on the structural complexity of the molecule. Descriptors not calculated for a given molecule were set to zero in subsequent QSPR model development. The descriptor set generated from CODESSA and Dragon was orthogonalized to remove repetitive and insignificant descriptors. This reduced set still contained hundreds of descriptors.

Using non-linear algorithms to find the best set of descriptors from hundreds of descriptors requires extensive computational time and is often impractical. Therefore, sequential multiple regression techniques were used to obtain a reduced set of descriptors. This technique entails the creation of the best QSPR model for a selected property based on the selection of a single descriptor. With the selected descriptor retained, this process is then repeated sequentially until the desired number of descriptors is realized. While this method is linear, the relationship between a descriptor and property of interest may often be non-linear. To ensure that *non-linear* relationships between descriptors and properties were not ignored, non-linear transformations of the descriptors were evaluated and an expanded set of descriptors were obtained before beginning the sequential regression analysis. The descriptor set was reduced to 40 descriptors using sequential regression analysis and further reduction was accomplished using the heuristic analysis available in CODESSA. The final set of descriptors was retained for non-linear regression using artificial neural networks.

Robust ANN algorithms have been developed which are capable of:

1. Finding the optimal network architecture: The number of hidden layers and the number of hidden neurons in these layers is varied to produce an optimal error function value.
2. Using cross-validation analysis to avoid over-fitting: The error on a separate validation set was monitored throughout the training process, and if the validation set error increases for six successive training iterations, training was stopped.
3. Dividing the data systematically into training, validation and testing sub-sets: Multiple random divisions of the entire data were performed and the data sets that result in the least error function values were recorded.
4. Employing effective normalization techniques: The initial input data were normalized to have zero mean and a standard deviation of 1. This ensures that abnormally small or large magnitude data do not skew the performance of the neural network.
5. Conducting multiple weight initializations: A typical error function surface has multiple minima and maxima. Therefore, multiple weight initializations were performed to maximize the likelihood of global minimization of the error function.
6. Utilizing multiple performance functions for analyzing the network: Using various statistical measures for the objective function during training allows for the best representation of the uncertainty present in the collected data of the property of interest.

The network performance was improved by studying networks with multiple transfer functions and numbers of neurons in the hidden layers. Network architectures with one or

two hidden layers have proven to be sufficient for non-linear regression and, hence, our algorithm searches for all possible one or two hidden layer architectures that satisfy a degree of freedom ratio (ratio of the number of network connections and the number of data points) lower limit of two (36). A small ensemble of networks was created using the average of the property values predicted by three independent networks. This technique serves to nullify the effects of biased networks.

3.2 CPE design

A database comprised of 160 human skin penetration coefficient measurements was used to develop a skin penetration QSPR model. Our QSPR model for skin penetration coefficient predicts the penetration data for the training, validation, and test sets considered with an absolute average percent deviation (%AAD) of 6, 14 and 13, respectively (8). Similarly, 2029 octanol/water partition coefficient data from the PhysProp database by Syracuse Research, 965 melting point data from the DIPPR database and roughly 900 skin sensitization data (10, 37) were used to develop the respective QSPR models. Data from local lymph node assay (LLNA) experimental procedure, guinea pig maximization test (GPMT) and Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) database were used to develop effective skin sensitization QSPR models. Since the experimental procedure and end-point ranking assigned to molecules by LLNA, GPMT and BgVV are different, three separate QSPR models were developed (10, 37). More details on the prediction networks used are provided in Table 1. Other properties, such as molecular weight, number of hydrogen-bond donors and number of hydrogen bond acceptors, were calculated using commercially available software.

As stated previously, 272 CPEs from the literature were used as input molecules for the first generation. Crossover and mutation operators were assigned equal probabilities of selection in the first generation and monitored in subsequent generations. Roughly 15 functional groups identified to be prominent in literature CPEs were used in functional group addition mutations. After each generation, the offspring molecules were initially monitored to remove any unrealistic molecular structure and large molecules. In this work, valency is used to determine feasibility of the new molecular structures. SMILES structure was used to generate the 2-D structure of the offspring molecules using Chem3DUltra software. The 3-D structures were generated and optimized for minimum energy. Molecular descriptors were generated using the previously mentioned software, and property predictions were made using robust QSPR models already developed. For each property within the acceptable range, the score of the molecule was incremented by a value of one. Thus, a summary numerical score is assigned to each of the molecules generated. Molecules that passed all the screening tests and had a fitness score value of eight were accepted as potential CPEs. Figure 5 summarizes the scoring of the molecules in each generation. Molecules that have high log K_p values have higher skin penetration ability. Hence the screened molecules are sorted according to their log K_p value and the top 10% of the molecules were retained and added to the parent molecule set to be used in the next generation. Approximately 1000 molecules were generated in each generation run, and this procedure was repeated for five generations. Table 2 summarizes the results of the five generation runs. In general, slightly less than 20% of the molecules generated were considered candidate for further evaluation as CPEs.

The molecules thus identified were further validated for skin permeation and skin irritation using carefully-designed experiments. The experimental work was performed by members of the OSU Chemical Design Research Group and is disseminated elsewhere (12-15); nevertheless, a brief discussion of the experimental work is provided for completeness. Molecules with a score of 8 and with high log K_p values in each generation were selected for experimental validation. In this work, insulin was the targeted drug to be delivered

transdermally; hence, the CPEs were experimentally validated for penetration enhancement of insulin. The skin permeation experimental procedures were validated by performing permeability measurements on four known CPEs. These validations used a Franz Cell with porcine skin and high-performance liquid chromatography (HPLC) analysis. The resistance factor (RF) and the insulin flux obtained using the Franz cell and the HPLC method, respectively, for the experimentally validated molecules are presented in Table 3. Chemical compounds with high RF and insulin flux values are considered effective in transdermal penetration.

The toxic effects of these enhancers were also studied on (a) human foreskin fibroblasts (HFFs) with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT)-formazan assays at two different concentrations, and (b) porcine abdominal skin using histology and haematoxylin/eosin (H/E) staining at the end of a 24-hour exposure period. A detailed discussion of the experimental procedures is beyond the scope of this paper and is given elsewhere (12-15).

Eighteen of the potential CPEs identified by the virtual design algorithm were validated experimentally for skin permeation and toxicity. As shown in Table 3, four molecules (OSU2, OSU3, OSU10 and OSU14) were found to be effective in enhancing insulin permeation through the skin with minimal or no toxic effects. Among the chemical classes identified to be effective in enhancing skin permeation were aldehydes, alkanes, alkenes, alcohols, acids and ketones. The physiochemical properties of the CPE molecules identified are shown in Table 3.

One of the major impediments to experimental validation of the newly generated chemicals is the lack of commercial availability of the chemicals. Often a chemical with good permeation and fitness scores is not available commercially, and its experimental validation was rendered more difficult. As such, in this study, we elected to validate experimentally only molecules that were available commercially, even though these molecules may not represent the highest fitness score in each generation. This limitation is further amplified by the fact that as the number of generations increases, the crossover and mutations among molecules becomes extensive, which, in turn, leads to greater numbers of novel molecules not available commercially.

Although a CPE may permeate through the skin effectively, its ability to enhance the permeation of a specific drug, such as insulin, through the skin depends, in part, on chemical interaction effects between the CPE and the drug. This explains why some of the virtually designed CPEs were not effective in transporting insulin through the skin. Knowledge of CPE-drug interactions in the pre-design stage is highly desired to enhance the predictive capability of our CAMD algorithms for transdermal drug delivery. From the current CPE CAMD algorithm, chemical compounds with hydrogen bonding groups and having log K_{ow} values greater than 2.5 were observed to be effective in enhancing insulin permeation through the skin. Acids, alcohols, aldehydes, and ketones are among the chemical classes found to be effective in enhancing insulin permeation through the skin. Incorporating this knowledge into future insulin CPE CAMD algorithms will further enhance the predictive capabilities of virtual design. Also, using known insulin CPEs as seed molecules can be effective in developing insulin-specific CPEs, as shown in our other work; see e.g. Yerramsetty et al. (15).

The new molecules evolved in each generation were subjected to a series of steps (e.g., conversion of the SMILES structure of the molecule to 2-D structure, conversion from 2-D to 3-D structure, optimization of the 3-D structure for minimum energy using Chem3DUltra, re-optimization of the molecule using AMPAC, generation of descriptors, property

prediction using the descriptors generated and scoring and screening based on the property values) before passing to the next generation. This process is laborious and becomes increasingly difficult as the number of generations increases. Further, the human involvement required currently hinders the ability of the GA design program to run multiple generations and limits the diversity in the population. Although some studies claim to have run multiple generations in their GA program, their discussions were limited concerning the optimization of the newly generated molecules. Hence, a need exists for an automation tool capable of handling multiple generations and achieving more genetic diversity with minimum manual intervention.

4. Conclusions

Genetic algorithm-based virtual design of molecules possessing desired properties offers rapid and reduced-cost development opportunities. Four novel molecules were identified to be effective in enhancing insulin skin permeation using virtual design algorithms and carefully designed experiments. Our results indicate that integrating genetic algorithms and non-linear QSPR modeling offers a reliable CAMD algorithm for generation of potential chemical penetration enhancers (CPEs). Further, these results demonstrate the efficacy of this virtual design approach in identifying potential CPEs for transdermal drug delivery of insulin.

The lack of accurate knowledge of the drug-chemical interactions in the pre-design stage is a limitation in the current methodology. The *a priori* knowledge of drug-chemical interactions would improve the design ability of our newly-developed algorithms and, thus, potentially reduce the number of experimental validations, which are often expensive and laborious.

A need exists for a computational platform to orchestrate the creation of multiple generations of CPE candidates with greater genetic diversity and minimum manual intervention. Further, synthesis of chemical compounds identified as effective CPEs would expand the list of insulin enhancers beyond the chemical structures available commercially, and this could potentially lead to identification of superior CPEs.

Acknowledgments

Financial support for this research was provided by the National Institutes of Health, the National Institute of Biomedical Imaging and Bioengineering (1R21EB005749).

References

1. Venkatasubramanian V, Chan K, Caruthers JM. Computer-aided molecular design using genetic algorithms. *Comput Chem Eng.* 1994; 18:833–44.
2. Sundaram A, Venkatasubramanian V. Parametric sensitivity and search-space characterization studies of genetic algorithms for computer-aided polymer design. *J Chem Info Comput Sci.* 1998; 38:1177–91.
3. Devillers, J. Genetic algorithms in molecular modeling. Academic Press; San Diego: 1996.
4. Harper PM, Gani R, Kolar P, Ishikawa T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilib.* 1999; 158:337–47.
5. Venkatasubramanian V, Chan K, Caruthers JM. Evolutionary design of molecules with desired properties using the genetic algorithm. *J Chem Inf Comput Sci.* 1995; 35:188–95.
6. Godavarthy, SS. Ph.D. Dissertation. School of Chemical Engineering, Oklahoma State University; Stillwater, Oklahoma: 2004. Design of improved solvents for extractive distillation.
7. Li J. CAMD in modern drug discovery. *Drug Discov Today.* 1996; 1:311–2.

8. Neely BJ, Madihally SV, Robinson RL Jr, Gasem KAM. Non-linear quantitative structure-property relationship modeling of skin permeation coefficient. *J Pharm Sci.* 2009; 98:4069–84. [PubMed: 19189399]
9. Godavarthy SS, Robinson RL Jr, Gasem KAM. An improved structure-property model for predicting melting-point temperatures. *Ind Eng Chem Res.* 2006; 45:5117–26.
10. Golla S, Madihally S, Robinson RL, Gasem KAM. Quantitative structure–property relationship modeling of skin sensitization: A quantitative prediction. *Toxicol in Vitro.* 2009; 23:454–65. [PubMed: 19162165]
11. Golla S, Madihally SV, Robinson RL Jr, Gasem KAM. Quantitative structure-property relationships modeling of skin irritation. *Toxicol in Vitro.* 2009; 23:176–84. [PubMed: 19027061]
12. Rachakonda VK, Yerramsetty KM, Madihally SV, Robinson RL Jr, Gasem KAM. Screening of chemical penetration enhancers for transdermal drug delivery using electrical resistance of skin. *Pharm Res.* 2008; 25:2697–704. [PubMed: 18683029]
13. Godavarthy SS, Yerramsetty KM, Rachakonda VK, Neely BJ, Madihally SV, Robinson RL Jr, et al. Design of improved permeation enhancers for transdermal drug delivery. *J Pharm Sci.* 2009; 98:4085–99. [PubMed: 19697392]
14. Yerramsetty KM, Neely BJ, Madihally SV, Gasem KAM. A skin permeability model of insulin in the presence of chemical penetration enhancer. *Int J Pharm.* 2010; 388:13–23. [PubMed: 20026200]
15. Yerramsetty KM, Rachakonda VK, Neely BJ, Madihally SV, Gasem KAM. Effect of different enhancers on the transdermal permeation of insulin analog. *Int J Pharm.* 2010; 398:83–92. [PubMed: 20667506]
16. Schult, CJ. Ph.D. Dissertation. School of Chemical Engineering, Oklahoma State University; Stillwater, Oklahoma: 2000. Design of solvents for extractive distillation.
17. Gasem, KAM.; Robinson, RL., Jr; Schult, CJ.; Todd, BA. Separation of hydrocarbons by extractive distillation. United States Patent No. 6392115. 2002.
18. Harper PM, Gani R. A multi-step and multi-level approach for computer aided molecular design. *Comput Chem Eng.* 2000; 24:677–83.
19. Constantinou L, Bagherpour K, Gani R, Klein JA, Wu DT. Computer aided product design: Problem formulations, methodology and applications. *Comput Chem Eng.* 1996; 20:685–702.
20. Achenie, LEK.; Gani, R.; Venkatasubramanian, V. Computer aided molecular design: Theory and practice. Elsevier Science; Amsterdam: 2003.
21. Finnin BC, Morgan TM. Transdermal penetration enhancers: Applications, limitations, and potential. *J Pharm Sci.* 1999; 88:955–8. [PubMed: 10514338]
22. Kumar R, Philip A. Modified transdermal technologies: Breaking the barriers of drug permeation via the skin. *Trop J Pharm Res.* 2007; 6:633–44.
23. Brown L, Langer R. Transdermal delivery of drugs. *Annu Rev Med.* 1988; 39:221–9. [PubMed: 3130771]
24. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2001; 46:3–26. [PubMed: 11259830]
25. Osborne DW, Henke JJ. Skin penetration enhancers cited in the technical literature. *Pharm Technol.* 1997; 21:58–66.
26. Thong HY, Zhai H, Maibach HI. Percutaneous penetration enhancers: An overview. *Skin Pharmacol Appl Skin Physiol.* 2007; 20:272–82.
27. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci.* 1988; 28:31–6.
28. Bone RG. SMILES extensions for pattern matching and molecular transformations: Applications in chemoinformatics. *J Chem Inf Comput Sci.* 1999; 39:846–60.
29. Douguet D, Thoreau E, Grassy G. A genetic algorithm for the automated generation of small organic molecules: Drug design using an evolutionary algorithm. *J Comput-Aided Mol Des.* 2000; 14:449–66. [PubMed: 10896317]
30. Morley, C. [accessed Dec 1, 2007] Openbabelgui. 2006. <http://openbabel.sourceforge.net/>

31. Jones, G. Genetic and evolutionary algorithms. John Wiley & Sons, Ltd; New York: 1998.
32. CambridgeSoft. Chembiooffice 11.0. Cambridge Software; 2008.
33. Semichem. Ampac, 6.0. Semichem Inc.; Shawnee, KS: 1998.
34. Semichem. Codessa, 2.7.8. Semichem Inc.; Shawnee, KS: 1998.
35. Dragon, 5.4. Milano Chemometrics; Milano, Italy: 2006.
36. Hagan, M. Personal communication. school of electrical and computer engineering, oklahoma state university; 2007.
37. Golla, S. MS Thesis. School of Chemical Engineering, Oklahoma State University; Stillwater, Oklahoma: 2008. Virtual design of chemical penetration enhancers.

Figure 1. Basic steps in CAMD

Figure 2. Virtual design of CPEs: Flow diagram

Figure 3. Crossover operators

Figure 4. Genetic operators

Figure 5. Scoring and screening of an offspring molecule

Table 1
Summary of QSPR models used for property prediction

Property	No. of data points used for modeling	Range of property values	No. of descriptors	Neural network architecture	R ²	RMSE ¹	% Accuracy ²
Melting point	965	14 to 586 K	20	20-14-7-1	0.90	25 K	-
Octanol/water partition coefficient (log K _{ow})	2029	-12 to 9.4	9	9-30-6-1	0.91	0.7	-
Skin penetration coefficient (log K _p)	160	-5.6 to -1.0	10	10-5-3-1	0.90	0.36	-
Skin sensitization							
LLNA	358	0 to 1	25	25-4-11-1	-	-	90
GPMT	307	0 to 1	25	25-3-6-1	-	-	95
B _g VV	251	0 to 1	24	24-4-1	-	-	93

¹ RMSE = average of root-mean-squared error in property predictions for training, validation and test sets

² % Accuracy = average of percentage of correct classifications for training, validation and test sets

Table 2
Genetic algorithm results from each generation

Generation number	Number of seed molecules	Total number of new molecules generated	Accepted molecules: score of 8	% of accepted molecules
1	249	943	120	12.7
2	269	978	155	15.9
3	290	995	269	27.0
4	311	1009	193	19.1
5	331	909	156	17.2
All	1450	4834	893	18.5

Table 3
Experimental results and predicted property values of CPEs

Generation	Chemical ¹	log K _{ow}	MW	nHacc	nHdon	MP	log K _p	BgVV Score	GPMT score	LLNA score	RF	Cumulative Insulin (10 ⁻⁴ gIU/m ²)
1	OSU1	2.1	120.0	1.0	1.0	291.1	-1.5	0.2	0.1	0.2	1.4 ± 0.1	1.8
	OSU2	2.8	128.0	1.0	1.0	242.9	-1.7	0.1	0.3	0.3	28 ± 7	8.1
	OSU3	2.8	128.0	1.0	1.0	264.2	-1.7	0.2	0.1	0.1	17 ± 2	5.5
	OSU4	2.7	72.1	0.0	0.0	124.2	-1.8	0.5	0.2	0.0	2.1 ± 0.1	1.9
	OSU5	1.7	100.0	1.0	1.0	200.4	-1.9	0.1	0.3	0.3	2.1 ± 0.3	4.5
	OSU6	2.0	130.0	0.0	0.0	199.0	-2.2	0.1	0.0	0.2	7 ± 1	6.8
	OSU7	2.4	70.1	0.0	0.0	121.8	-2.2	0.1	0.1	0.1	7 ± 3	-
	OSU8	1.0	86.1	1.0	1.0	193.4	-2.2	0.1	0.3	0.3	-	2.6
2	OSU9	2.2	144.0	0.0	0.0	208.4	-2.5	0.1	0.1	0.2	3 ± 1	4.5
	OSU10	2.9	168.3	1.0	1.0	233.1	-1.6	0.2	0.0	0.3	53 ± 6	10.5
	OSU11	1.3	116.2	1.0	1.0	264.9	-2.3	0.1	0.1	0.0	5 ± 2	2.0
	OSU12	1.1	87.2	1.0	1.0	131.5	-2.6	0.5	0.0	0.1	60 ± 8	6.6
3	OSU13	1.6	116.2	0.0	0.0	175.8	-2.0	0.2	0.1	0.0	2 ± 1	3.0
	OSU14	1.8	100.2	1.0	1.0	187.7	-2.1	0.1	0.1	0.3	14 ± 2	3.8
	OSU15	2.5	129.2	1.0	1.0	262.2	-2.2	0.2	0.2	0.2	76 ± 8	10.3
4	OSU16	2.1	114.2	1.0	1.0	221.8	-1.5	0.2	0.1	0.1	4 ± 2	3.1
	OSU17	1.5	112.2	1.0	1.0	203.4	-1.9	0.2	0.4	0.2	4 ± 2	3.4
5	OSU18	2.3	114.2	1.0	1.0	254.1	-1.4	0.1	0.1	0.1	5.0 ± 0.6	-

¹The chemical names of the identified CPE molecules have been withheld to protect intellectual property rights