molecular pharmaceutics

ACS AuthorChoice

Article

pubs.acs.org/molecularpharmaceutics

Terms of Use

Computational Predictions of Glass-Forming Ability and Crystallization Tendency of Drug Molecules

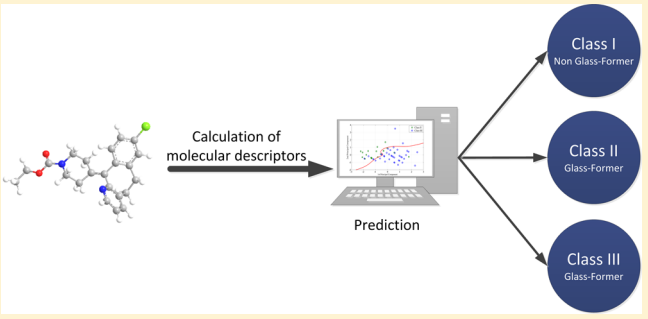
Amjad Alhalaweh,[†] Ahmad Alzghoul,[‡] Waseem Kaialy,[§] Denny Mahlin,[†] and Christel A. S. Bergström*[,†]

[†]Department of Pharmacy, Uppsala University, Uppsala Biomedical Centre, P.O. Box 580, SE-751 23 Uppsala, Sweden

[‡]Department of Information Technology, Uppsala University, Lägerhyddsv. 2, hus 1, Box 337, SE-751 05 Uppsala, Sweden

[§]School of Pharmacy, Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton WV1 1LY, United Kingdom

ABSTRACT: Amorphization is an attractive formulation technique for drugs suffering from poor aqueous solubility as a result of their high lattice energy. Computational models that can predict the material properties associated with amorphization, such as glass-forming ability (GFA) and crystallization behavior in the dry state, would be a time-saving, cost-effective, and material-sparing approach compared to traditional experimental procedures. This article presents predictive models of these properties developed using support vector machine (SVM) algorithm. The GFA and crystallization tendency were investigated by melt-quenching 131 drug molecules *in situ* using differential scanning calorimetry. The SVM algorithm was used to develop computational models based on calculated molecular descriptors. The analyses confirmed the previously suggested cutoff molecular weight (MW) of 300 for glass-formers, and also clarified the extent to which MW can be used to predict the GFA of compounds with MW < 300. The topological equivalent of Grav3_3D, which is related to molecular size and shape, was a better descriptor than MW for GFA; it was able to accurately predict 86% of the data set regardless of MW. The potential for crystallization was predicted using molecular descriptors reflecting Hückel pi atomic charges and the number of hydrogen bond acceptors. The models developed could be used in the early drug development stage to indicate whether amorphization would be a suitable formulation strategy for improving the dissolution and/or apparent solubility of poorly soluble compounds.



KEYWORDS: *amorphous, glass forming ability, crystallization tendency, support vector machine, molecular descriptors*

INTRODUCTION

Drug compounds can exist in different solid forms. On solidification, single-component systems can assume a crystalline form with long- and short-range order, some of which display polymorphism, or a glass form with only short-range order. Multicomponent systems may also exist in several different solid forms, and in these, long-range order is obtained by ordered packing of two or more components. Examples of such systems include crystalline forms of drug molecules in combination with counterions (salts), solvents (solvates/hydrates), or inert guest molecules (cocrystals).[1−3] A multicomponent system can also exist in the amorphous form, either in the recently established coamorphous state in which a drug and a guest molecule interact to form an amorphous material or in the traditionally used amorphous solid dispersions using polymer matrices.[4,5] Screening for the type of solid system that should be used to improve the performance of a drug has been the subject of great interest.[2,6] The aim of these and other screening exercises is to avoid the risk of using metastable forms, to comply with regulatory guidelines, or to identify potential strategies for improving delivery of poorly water-soluble compounds so as to transform them into functional medicines.[6−9]

The amorphous form of a material (also known as a glass) is formed by processing, and thus disordering, the crystalline form. Amorphous drugs have been extensively investigated as this state has the potential to overcome issues related to dissolution rate and/or solubility-limited absorption. This strategy is of particular interest for nonionizable compounds, as the commonly used salt formation cannot be applied to such compounds.[10,11] The number of marketed dosage forms taking advantage of amorphization is growing and is likely to increase further as a response to the limited solubility often associated with new molecules. Prediction of glass formation using a theoretical, computational approach is therefore warranted to reduce time, cost, and material demands when deciding upon formulation strategies to be explored experimentally in the laboratory.[12−14]

Received: April 24, 2014
Revised: July 9, 2014
Accepted: July 11, 2014
Published: July 11, 2014

The glass-forming ability (GFA) of a material has been defined as the ability of the material to vitrify on cooling from the melted state or to form a glass as a result of processes including spray-drying or mechanical activation.[15] Predictions of the GFA, based on measured parameters, have been published previously for various materials.[12−14,16−18] Previous investigations have indicated that the Hruby parameter strongly correlates with the stability of the glass and the glass-forming tendency for inorganic materials.[18] Researchers have studied the GFA of drugs and related this property to both calculated molecular descriptors and experimentally determined properties. Molecular weight (MW) has been found to be an important parameter for determining the GFA.[12,13] Using a data set of 50 compounds, we have previously identified that drug molecules with a MW above 300 g/mol are glass-formers.[13] This hypothesis was then confirmed using a test set of an additional 50 compounds.[13] However, for compounds with a MW below 300 g/mol, no clear trends were observed for this relatively large data set. Other studies have indicated that the molecular flexibility (reflected by the number of rotatable bonds) is another important property for predicting GFA, with flexible compounds being more likely to form the amorphous state.[12] In addition to MW and rotatable bonds, branched carbon skeletons, the presence of electronegative atoms, and a low number of aromatic rings have also been suggested to drive GFA.[14]

A recent study classified the crystallization tendencies for the amorphous forms of drug molecules according to their crystallization behavior from the undercooled melt.[12] In that study a data set of 51 drugs was used and compounds that crystallized on cooling the undercooled melt at a temperature lower than the melting temperature were considered to be nonglass-formers (nGFs), defined as Class I compounds (Figure 1a). Glass-formers (GFs), which crystallized on heating the melt-quenched material above the T_g, were defined as Class II compounds (Figure 1b), and compounds that showed no sign of crystallization on heating the melt-quenched material were classified as Class III compounds (Figure 1c). This analysis of the crystallization tendency needs to be experimentally assessed and computational models for prediction are therefore warranted.

In this study, the molecular properties considered important for predicting the GFA and the crystallization tendency were explored using a data set of 131 compounds. The following specific objectives were explored, using this large data set: (1) whether the MW can also predict the GFA for compounds with MW < 300; (2) whether other molecular descriptors can predict the GFA better than the MW; and (3) whether the crystallization tendency can be predicted using calculated molecular descriptors.

EXPERIMENTAL SECTION

Materials. All chemicals were purchased from standard suppliers and were of high purity (98−99%). The compounds were selected to be structurally diverse (Table 1) to enable the application of the developed models to a wide range of drug molecules.

Production of the Amorphous State. The compounds were amorphized by *in situ* quenching in a differential scanning calorimetry (DSC) Q2000 calorimeter (TA Instruments), which was calibrated for temperature and enthalpy using indium. The instrument was equipped with a refrigerated cooling system. The melting point was determined using 1−3 mg of each compound in nonhermetic aluminum pans. The compounds were scanned at a heating rate of 10 °C/min under a continuously purged dry nitrogen atmosphere (50 mL/min).

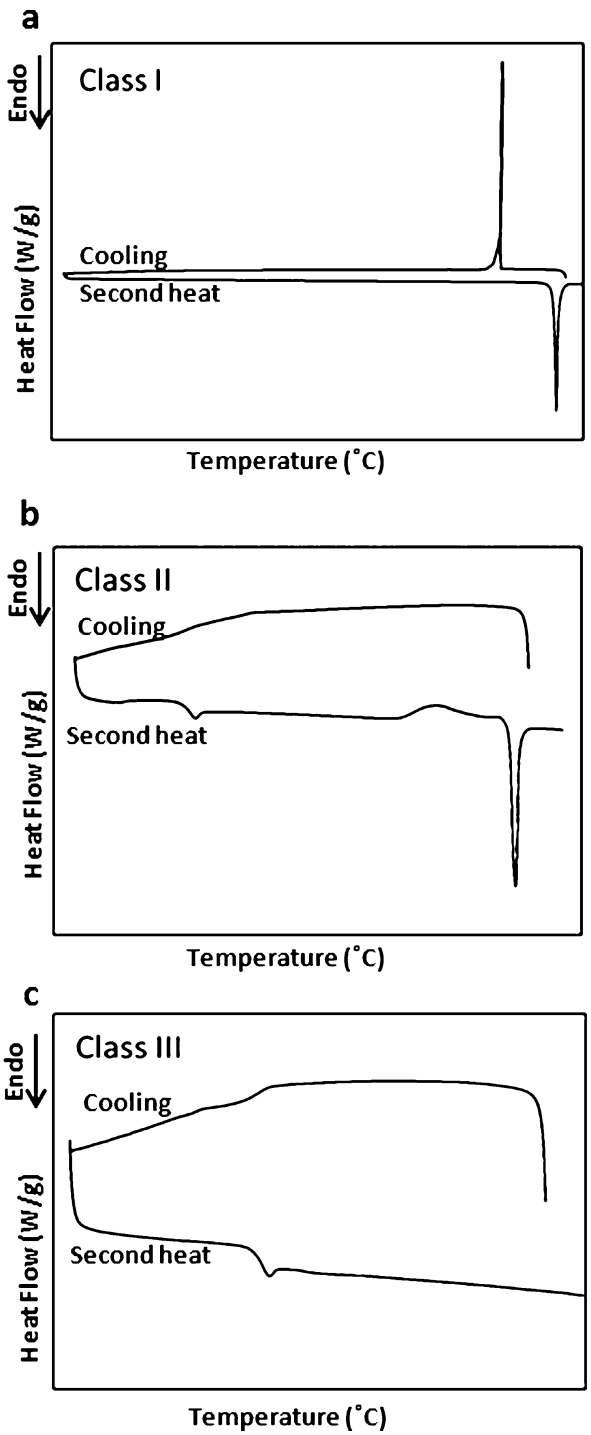


Figure 1. Differential scanning calorimetry thermograms showing the crystallization behavior of the three classes. (a) Class I crystallization upon cooling. (b) Class II crystallization in the second heat. (c) Class III, no crystallization was observed.

Glass formation was investigated by weighing 1−3 mg of the compound into nonhermetic pans and heating to around 2 °C above the peak melting point using an equilibrate function. The system was kept isothermal at this temperature for 2 min to ensure complete melting and was thereafter cooled to −70 °C at a ramp rate of 20 °C/min. The formation of a glass state was

dx.doi.org/10.1021/mp500303a | *Mol. Pharmaceutics* 2014, 11, 3123−3132

Table 1. Glass-Forming Ability and Calculated Molecular Properties for the Data Set Studied; Compounds Are Organized Based on Alphabetical Order within the Class with Training Set Compounds Listed First[a]

compound	class	MW	amide	ring	pi Fukui	T_Grav3	pi atomic	HBA	log P	rot bonds	Tr/Te
aceclofenac	III	354.2	0	0	8.3	16.7	0.94	5	2.8	7	Tr
acemetacin	III	415.8	1	1	14.1	17.7	1.34	6	2.0	6	Tr
albendazole	III	265.3	0	1	12.5	15.7	0.57	4	1.9	6	Tr
bezafibrate	III	361.8	1	0	12.6	16.8	0.77	5	3.0	7	Tr
bicalutamide	III	430.4	1	0	27.5	18.6	0.82	6	1.9	7	Tr
captopril	III	217.3	1	1	2.6	14.2	0.46	4	0.2	4	Tr
clotrimazole	III	344.9	0	1	16.9	17.0	0.29	1	4.5	4	Tr
diatrizoic acid	III	613.9	2	0	4.1	19.5	0.97	6	2.4	4	Tr
diazepam	III	284.8	1	1	9.8	15.8	0.55	3	3.4	0	Tr
d-lactose	III	342.3	0	1	0.0	16.8	0.00	11	−3.9	4	Tr
ezetimibe	III	409.4	1	1	26.3	17.9	0.74	4	4.9	6	Tr
famotidine	III	337.5	0	1	14.4	18.0	0.53	9	−0.2	6	Tr
felodipine	III	384.3	0	1	25.4	17.1	0.93	5	2.4	4	Tr
fenofibrate	III	360.8	0	0	9.4	16.8	0.91	4	3.9	5	Tr
glafenine	III	372.8	0	1	15.4	17.2	0.90	6	2.1	6	Tr
glibenclamide	III	494.0	2	1	10.5	19.7	0.81	8	3.4	10	Tr
hydrochlorothiazide	III	297.7	0	1	2.1	17.9	0.61	7	−0.6	1	Tr
hydrocortisone	III	362.5	0	1	1.7	17.2	0.69	5	1.6	2	Tr
hydroflumethiazide	III	331.3	0	1	8.1	18.3	0.48	7	−0.1	2	Tr
ibuprofen	III	206.3	0	0	5.9	13.8	0.30	2	2.8	4	Tr
isradipine	III	371.4	0	1	43.5	17.2	0.86	8	0.8	4	Tr
itraconazole	III	705.7	1	1	29.5	21.8	1.02	9	4.2	10	Tr
ketoconazole	III	531.4	1	1	15.5	19.7	0.90	7	2.6	8	Tr
ketoprofen	III	254.3	0	0	23.2	15.0	0.72	3	3.0	2	Tr
linaprazan	III	366.5	1	1	28.8	17.1	0.73	5	2.2	6	Tr
loratadine	III	382.9	0	1	9.3	17.6	0.49	4	3.9	3	Tr
metolazone	III	365.8	1	1	9.6	18.0	0.87	6	2.1	2	Tr
miconazole	III	416.1	0	1	9.7	17.8	0.46	2	4.5	6	Tr
nilutamide	III	317.2	2	1	39.8	16.1	0.65	6	2.6	2	Tr
nimesulide	III	308.3	0	0	31.3	17.2	0.43	6	2.6	4	Tr
nizatidine	III	331.5	0	1	15.2	17.0	0.50	6	0.9	9	Tr
omeprazole	III	345.4	0	1	22.7	17.5	0.59	5	0.6	5	Tr
pimozide	III	461.6	0	1	25.5	18.8	0.53	2	5.4	7	Tr
prednisone	III	358.4	0	1	5.2	17.1	1.05	5	1.4	2	Tr
probucol	III	516.9	0	0	56.1	19.0	0.46	2	6.8	8	Tr
simvastatin	III	418.8	0	1	4.3	18.0	0.51	5	3.2	7	Tr
spironolactone	III	416.6	0	1	3.5	18.4	0.90	4	34	2	Tr
sulfamerazine	III	264.3	0	1	19.0	16.4	0.81	6	0.8	3	Tr
sulfathiazole	III	255.3	0	1	16.9	16.6	0.67	5	0.2	3	Tr
warfarin	III	308.3	0	1	43.0	16.1	1.03	4	23	4	Tr
zolmitriptan	III	287.4	0	1	12.4	16.0	0.56	4	1.4	5	Tr
bucindolol	III	363.5	0	1	23.8	17.1	0.79	4	1.9	8	Te
budesonide	III	430.5	0	1	5.2	18.4	0.73	6	2.0	4	Te
carvedilol	III	406.5	0	1	31.4	18.0	0.83	5	1.7	10	Te
chloramphenicol	III	323.1	1	0	20.6	16.1	0.39	6	1.2	6	Te
chlorhexidine	III	505.5	0	0	11.8	18.8	0.86	10	4.8	11	Te
emtricitabine	III	247.2	0	1	7.0	15.4	0.41	5	0.4	1	Te
physostigmine	III	275.4	0	1	16.1	15.8	0.47	5	2.3	3	Te
probucol	III	516.9	0	0	56.1	19.0	0.46	2	6.8	8	Te
ritonavir	III	721.0	1	1	23.1	21.8	0.98	11	4.6	22	Te
sulindac	III	356.4	0	1	27.2	17.5	0.74	3	3.5	3	Te
acetaminophen	II	151.2	1	0	10.1	12.4	0.48	3	1.1	2	Tr
acetohexamide	II	324.4	1	1	10.3	17.4	0.64	6	1.7	5	Tr
bifonazole	II	310.4	0	1	32.6	16.5	0.22	1	4.0	3	Tr
celecoxib	II	381.4	0	1	19.2	18.2	0.45	4	3.2	2	Tr
cimetidine	II	252.3	0	1	11.6	15.4	0.40	5	−0.4	8	Tr
cinnarizine	II	368.5	0	1	21.5	17.5	0.03	2	4.5	5	Tr
clemastine	II	343.9	0	1	6.1	16.9	0.09	2	4.3	6	Tr
clofoctol	II	365.4	0	0	10.2	16.7	0.31	1	6.3	5	Tr
codeine	II	299.4	0	1	6.0	16.6	0.29	4	2.2	1	Tr

dx.doi.org/10.1021/mp500303a | Mol. Pharmaceutics 2014, 11, 3123−3132

Table 1. continued

compound	class	MW	amide	ring	pi Fukui	T_Grav3	pi atomic	HBA	log P	rot bonds	Tr/Te
danazol	II	337.5	0	1	9.4	17.1	0.15	3	4.2	1	Tr
dibucaine	II	343.5	1	1	13.5	16.8	0.41	5	2.5	10	Tr
d-salicin	II	286.3	0	1	8.1	15.8	0.20	7	−1.2	4	Tr
estradiol	II	272.4	0	1	8.1	15.9	0.22	2	3.6	0	Tr
flurbiprofen	II	244.3	0	0	13.1	14.8	0.38	2	3.5	2	Tr
nandrolone	II	274.4	0	1	1.7	16.0	0.41	2	3.5	0	Tr
nifedipine	II	346.3	0	1	43.5	16.6	1.00	7	1.3	3	Tr
testosterone	II	288.4	0	1	1.7	16.2	0.40	2	3.7	0	Tr
tinidazole	II	247.3	0	1	12.8	16.1	0.20	5	−0.2	4	Tr
tolazamide	II	311.4	0	1	7.0	17.3	0.40	6	2.0	5	Tr
aprepitant	II	534.4	1	1	17.9	19.5	0.55	5	4.7	8	Te
aripiprazole	II	448.4	1	1	15.5	18.5	0.94	5	3.6	7	Te
fluorescamine	II	278.3	0	1	66.9	15.9	0.83	4	2.3	0	Te
orlistat	II	495.8	1	1	3.8	18.9	0.72	6	3.7	1	Te
tamoxifen	II	371.5	0	0	49.2	17.3	0.24	2	5.2	5	Te
tenofovir	II	287.2	0	1	12.8	17.0	0.29	8	−1.1	5	Te
piroxicam	II	331.4	1	1	32.0	17.6	0.75	7	1.1	2	Tr
4-aminobenzoic acid	I	137.1	0	0	15.8	11.9	0.54	3	−0.4	0	Tr
aspirin	I	180.2	0	0	24.7	13.1	0.76	4	1.4	2	Tr
atenolol	I	266.3	1	0	9.3	15.2	0.44	5	0.9	8	Tr
benzocaine	I	165.2	0	0	15.7	12.9	0.53	3	1.4	2	Tr
bucetin	I	223.3	1	0	9.4	14.3	0.49	4	1.5	6	Tr
bufexamac	I	223.3	0	0	9.3	14.3	0.44	4	2.0	7	Tr
caffeine	I	194.2	2	1	19.9	13.9	0.50	3	0.1	0	Tr
carbamazepine	I	236.3	0	1	49.0	14.9	0.56	3	3.1	1	Tr
chlorpropamide	I	276.7	1	0	3.9	16.5	0.40	5	1.8	6	Tr
chlorzoxazone	I	169.6	0	1	3.9	13.2	0.41	2	1.3	0	Tr
diflunisal	I	250.2	0	0	21.9	14.8	0.72	3	3.6	0	Tr
d-mannitol	I	182.2	0	0	0.0	13.1	0.00	6	−2.5	5	Tr
fenbufen	I	254.3	0	0	25.4	15.0	0.69	3	2.7	3	Tr
flufenamic acid	I	281.2	0	0	46.3	15.4	0.69	3	2.9	3	Tr
flutamide	I	276.2	1	0	27.1	15.2	0.50	4	3.2	4	Tr
haloperidol	I	375.9	0	1	9.3	17.3	0.57	3	4.0	5	Tr
hymecromone	I	176.2	0	1	17.5	13.3	0.81	3	1.1	0	Tr
indoprofen	I	281.3	1	1	43.2	15.7	0.74	4	2.4	3	Tr
lidocaine	I	234.3	1	0	15.3	14.5	0.38	3	2.5	6	Tr
mefenamic acid	I	241.3	0	0	46.1	14.7	0.70	3	2.5	2	Tr
naproxen	I	230.3	0	0	19.9	14.5	0.49	3	2.4	3	Tr
pentachlorophenol	I	266.3	0	0	1.7	14.7	0.38	1	3.9	0	Tr
phenacetin	I	179.2	1	0	9.4	13.3	0.49	3	1.7	4	Tr
phenytoin	I	252.3	2	1	14.6	15.2	0.42	4	2.2	2	Tr
pindolol	I	248.3	0	1	11.7	15.1	0.58	3	0.8	6	Tr
primidone	I	218.3	2	1	13.2	14.4	0.46	4	1.2	2	Tr
probenecid	I	285.4	0	0	10.5	16.6	0.39	5	1.6	6	Tr
pyrazinecarboxamide	I	123.1	1	1	10.2	11.6	0.46	4	−1.0	0	Tr
saccharin	I	183.2	0	1	10.7	15.0	0.37	4	0.2	0	Tr
salicylic acid	I	138.1	0	0	15.6	12.0	0.52	3	1.2	0	Tr
spiperone	I	395.5	1	1	29.1	17.8	0.97	5	3.4	5	Tr
sulfadiazine	I	250.3	0	1	19.1	16.2	0.79	6	0.5	3	Tr
sulfamethoxazole	I	253.3	0	1	16.7	16.3	0.64	6	0.6	3	Tr
sulfanilamide	I	172.2	0	0	8.3	14.6	0.39	4	−0.1	1	Tr
theophylline	I	180.2	2	1	20.1	13.6	0.51	3	−0.3	0	Tr
tolfenamic acid	I	261.7	0	0	9.4	15.1	0.73	3	2.7	2	Tr
trimethoprim	I	290.3	0	1	17.2	15.8	0.82	7	0.9	5	Tr
tyramine	I	137.2	0	0	8.0	12.1	0.19	2	1.3	2	Tr
zoxazolamine	I	168.6	0	1	3.9	13.2	0.37	3	1.4	0	Tr
antipyrine	I	188.2	0	1	14.6	13.7	0.56	1	2.2	0	Te
artemisinin	I	282.3	0	1	1.3	16.2	0.26	5	2.4	0	Te
benzanthracene	I	228.3	0	0	30.1	14.9	0.00	0	5.1	0	Te
chrysin	I	254.2	0	1	47.0	15.2	1.09	4	2.1	0	Te
diethylstilbestrol	I	268.4	0	0	16.1	15.2	0.41	2	4.0	2	Te

dx.doi.org/10.1021/mp500303a | *Mol. Pharmaceutics* 2014, 11, 3123−3132

Table 1. continued

compound	class	MW	amide	ring	pi Fukui	T_Grav3	pi atomic	HBA	log P	rot bonds	Tr/Te
diuron	I	233.1	0	0	2.1	14.3	0.43	3	3.1	3	Te
felbinac	I	212.3	0	0	19.9	14.2	0.31	2	2.8	2	Te
flumequine	I	261.3	0	1	17.7	15.3	0.74	3	1.1	0	Te
furafylline	I	260.3	2	1	29.8	15.5	0.84	4	0.4	2	Te
gemfibrozil	I	250.3	0	0	9.3	14.7	0.48	3	2.8	6	Te
griseofulvin	I	352.8	0	1	7.4	16.9	1.34	6	1.0	3	Te
niclosamide	I	327.1	1	0	7.0	16.2	0.73	5	3.6	2	Te
nicotinamide	I	122.1	1	1	10.0	11.6	0.47	3	−0.3	0	Te
phenacetin	I	179.2	1	0	9.4	13.3	0.49	3	1.7	4	Te
urea	I	60.1	0	0	4.0	8.3	0.18	3	−1.5	0	Te

[a]The following abbreviations are used: crystallization tendency class I–III (class); molecular weight (MW); amide, number of amide groups; ring, indicator variable for the presence of ring structures except benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings); pi Fukui, second component of the autocorrelation vector of pi Fukui(−) indices; T_Grav3, topological equivalent of Grav3_3D; pi atomic, sum of absolute values of Hückel pi atomic charges but only on C atoms; HBA, number of hydrogen bond acceptors; log P; rot bonds, number of freely rotatable bonds; Tr/Te, training/test.

then investigated by heating the system again, at a heating rate of 20 °C/min, immediately after cooling. The presence of the amorphous form was indicated by detection of the T_g upon heating.[12]

Model Development Using Support Vector Machines. Molecular descriptors were calculated with the software ADMET Predictor (SimulationsPlus, CA) using molecules represented as structure-data files (sdf). In total, 284 descriptors were used as input for the modeling.

Support vector machines (SVMs) were used to develop the GFA and crystallization tendency models. SVMs are supervised learning models, which can be used to solve classification and regression problems. They work by representing the data in a higher dimensional space than that of the original. The SVM algorithm is used to find a hyper-plane that maximizes the margin between the two classes, as shown in Figure 2.[19] The

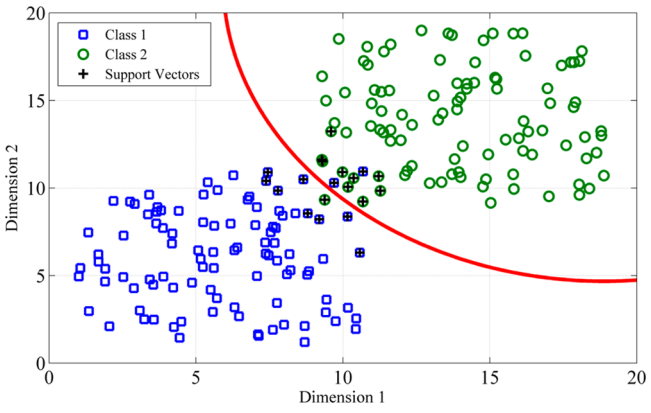


Figure 2. Classification of hypothetical data by support vector machines. If the test data fall on the right-hand side of the decision boundary, the data will be classified as Class 2, whereas if the data fall on the left side of the decision boundary, it will be classified as Class 1.

hyper-plane, also called the decision boundary, is defined by the support vectors. SVMs can also take advantage of nonlinear kernels such as polynomial and Gaussian functions to map the data to a high dimensional space where it can be linearly separated. Further, SVMs can handle nonseparable data sets by using the soft margin approach. The soft margin approach allows the SVM algorithm to define a decision boundary that tolerates small training errors. The soft margin approach allows the decision boundary to have a wider margin and thus be less susceptible to overfitting.

Various kernels were tested in this study, and the one with the best performance was selected. Principal component analysis (PCA) was used to map the multidimensional data into a two-dimensional space. The two dimensions refer to the first and the second principal components. The first principal component has the largest possible variance, and the second principal component has the second largest possible variance in the data set. We expect that variables with high impact on the output (GFA or crystallization tendency) will produce distinct areas of the two classes (i.e., GF and nGF). The classification accuracy of the SVM algorithm was used as a measure of the accuracy of these distinct areas. Thus, the subset of optimal variables was selected based on the classification accuracy of the SVM. Further details of the model development are described below.

A set of 100 compounds was used to provide training data in the study; the GFA values for these compounds were taken from the literature[12,13] or obtained in our laboratory using the same techniques used in the literature. A test set of 31 compounds was used to validate the models. The model development was performed as follows (Figure 3):

Step 1: A two-sample t test was applied to the training data and variables with a p value less than 0.01 were identified.

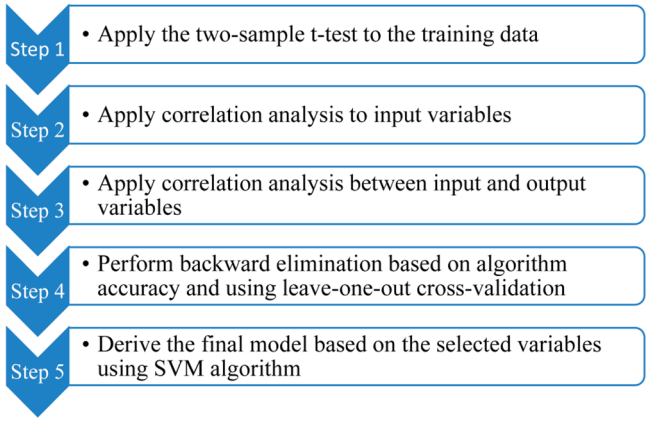


Figure 3. Protocol used for the development of the models.

dx.doi.org/10.1021/mp500303a | Mol. Pharmaceutics 2014, 11, 3123−3132

Step 2: A correlation analysis of the variables identified in Step 1 was carried out. Highly correlated variables were then clustered into the same group.

Step 3: The linear relationship between the input and output variables was investigated. For every group found in Step 2, the variable with the highest correlation with the output was selected to represent the group in the next steps. The number of variables was therefore reduced by Steps 2 and 3.

Step 4: A backward elimination method was implemented on the variables selected in Step 3 to remove those with the least impact on the prediction. PCA was used to map the high-dimensional data into two dimensions. Subsequently, SVMs were used to find the decision boundary separating the two classes. The leave-one-out (LOO) cross-validation technique was used on the training data to assess the performance of the SVMs for the given variables.

Step 5: A model was developed using the selected variables. The variables remaining in Step 4 were used to construct the decision boundary based on the SVM algorithm. In addition, a decision tree based on the selected variables was built.

Steps 1−5 were followed for two different scenarios:

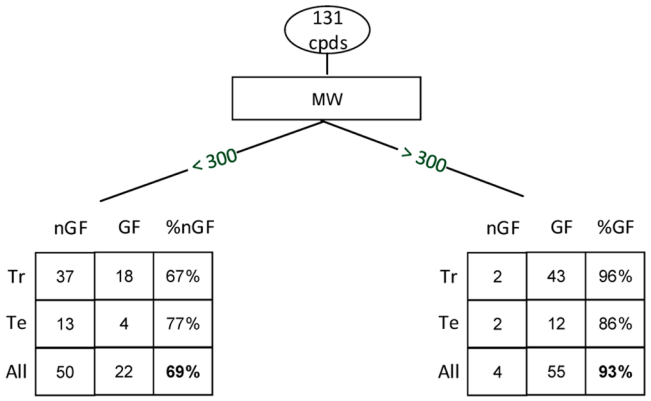
Case 1. Prediction of GFA using MW as a descriptor; in this case, the MW was used and its highly correlated variables were eliminated. Analysis then proceeded as described above.

Case 2. Prediction of GFA with no descriptor preference; in this case, all 284 variables were used to develop the model, without preference for the MW.

Prediction of the Crystallization Tendency. Descriptors distinguishing between Class II and III compounds were identified using the same strategy as outlined in Case 2 above.

RESULTS AND DISCUSSION

Experimental Procedures and Assessment of GFA. The GFA and crystallization tendencies of the compounds under study are presented in Table 1. The final list contains 131 compounds. While several others were studied with this approach, it was observed that these degraded on melting and that it was not possible to include them in the final data set.



	nGF	GF	%nGF
Tr	37	18	67%
Te	13	4	77%
All	50	22	**69%**

	nGF	GF	%GF
Tr	2	43	96%
Te	2	12	86%
All	4	55	**93%**

Figure 4. Prediction of glass-forming ability using molecular weight (MW) as the only descriptor. Abbreviations used: compounds (cpds); glass-former (GF); nonglass-former (nGF); training set (Tr); and test set (Te).

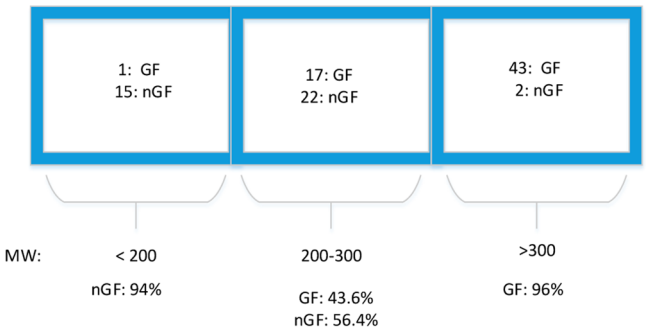


Figure 5. Experimental results according to their molecular weight (MW) and glass-forming ability for the training data set. Abbreviations used: glass-former (GF) and nonglass-former (nGF).

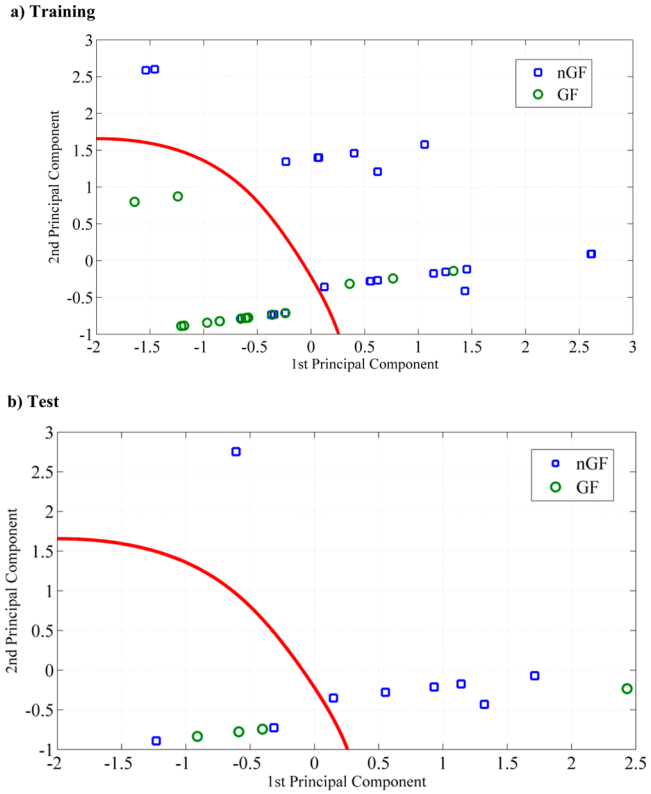


Figure 6. Prediction of the glass-forming ability using the support vector machine algorithm for compounds with molecular weights in the range of 200−300 g/mol: (a) training and (b) test. Abbreviations used: glass-former (GF) and nonglass-former (nGF).

The cooling rate used to vitrify the compounds during the DSC experiment is an important factor that affects the crystallization of a compound but because the same experimental conditions were applied to all the drugs, it was felt that their inherent properties were accurately reflected. The assessment of GFA and crystallization tendency was adopted from Baird et al. (Figure 1).[12] This method was found to be robust and rapidly assessed both GFA and crystallization tendency and hence made the studies of a large number of compounds possible.

Prediction of Glass-Forming Ability Using Molecular Weight as Single Descriptor. The MW has been suggested previously as a suitable predictor for GFA.[12,13] GF compounds with a MW larger than 300 g/mol were well predicted but nGF compounds were poorly predicted, with an accuracy of 69% (Figure 4). The relationship between the GFA and MW was

dx.doi.org/10.1021/mp500303a | Mol. Pharmaceutics 2014, 11, 3123−3132

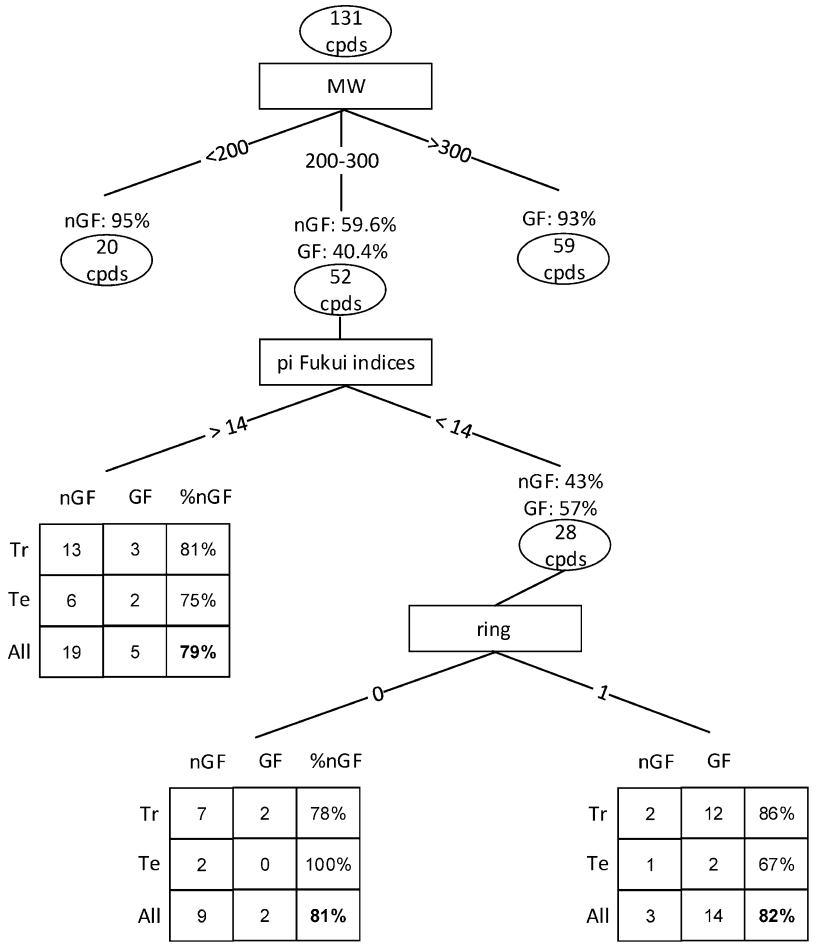


Figure 7. Molecular descriptors predictive for glass-forming ability suggested by the model developed in this study. Abbreviations used: compounds (cpds); glass-former (GF); nonglass-former (nGF); molecular weight (MW); training set (Tr); and test set (Te).

therefore further investigated in the MW range below 300 g/mol (Figure 5). Compounds with MW less than 200 g/mol were predicted to be nGF with an accuracy of 94%, but for compounds with MW in the range 200 to 300 g/mol, GFA was poorly related to MW.

To find out if other molecular descriptors would relate to GFA better for compounds with MW in the range 200−300 g/mol, SVMs were employed to analyze the relationships for 284 descriptors to GFA. A wide range of chemical descriptors, including topological, spatial, electrostatic, thermodynamic, and structural descriptors, among others, was used. The analysis indicated that three descriptors, reflecting pi Fukui indices, the presence of ring structures except for benzene and its condensed rings, and the number of amide groups, predicted the GFA well in both the training (82%) and test (77%) data sets (Figure 6). The results showed that a high value for the second component of the autocorrelation vector of pi Fukui indices indicated that the compound was an nGF. In contrast, the presence of ring structures except for benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings) and an increased number of amide groups indicated GFA. These findings were simplified to develop guidelines for compounds with MWs in the range 200−300 g/mol without the need of the SVM model (Figure 7). A decision tree based on pi Fukui indices and ring structures was subsequently found to predict GFA with high accuracy, without the use of SVM and the amide descriptor. Importantly, the developed models and rule-based systems all predicted the test set with high accuracy. The Fukui index is a reactivity descriptor used in conceptual density functional theory (DFT).[20] It indicates the tendency of an atom to lose or accept an electron (i.e., to undergo a nucleophilic or electrophilic attack). The Fukui indices thus describe the ability of molecules to become polarized when changes to the electron density occur. The Fukui function is affected by the number of molecules in a unit cell and the size of the unit cell. This function has been used to study the origin of polymorphic formation.[21] It is reported that high crystallization forces, reflected by increased effective electron reactivity (transfer and sharing), are associated with high values for the Fukui indices. These results are in agreement with our findings, where high Fukui index values were associated with high crystallization tendencies. The ring structure of the molecule is the driving force for GFA except in the case of benzene and its condensed rings. Clearly, aromatic ring structures are more rigid than aliphatic rings, and this typically leads to denser packing in the crystal lattice, with a subsequent increased crystallization tendency. In addition, the benzene structure has a relatively high electron reactivity, which, like the Fukui indices, could further push the compounds to be nGFs.[20]

Of the 20 compounds that had a MW < 200 g/mol, only acetaminophen was found to be a GF (Figure 7). Of the 59 compounds that were classified as nGFs, griseofulvin, haloperidol, niclosamide, and spiperone had MWs > 300 g/mol and hence violated the rule that compounds with MW >

dx.doi.org/10.1021/mp500303a | Mol. Pharmaceutics 2014, 11, 3123−3132

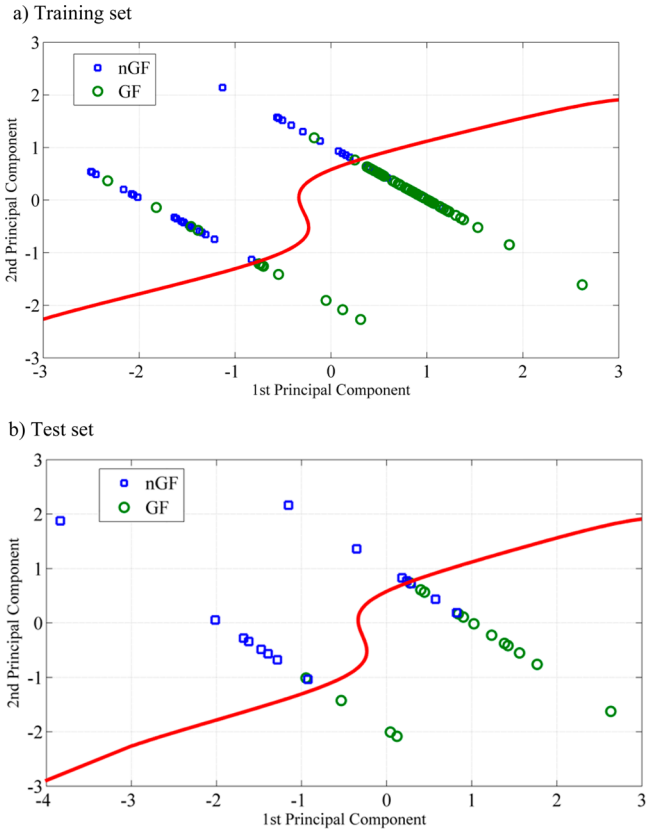
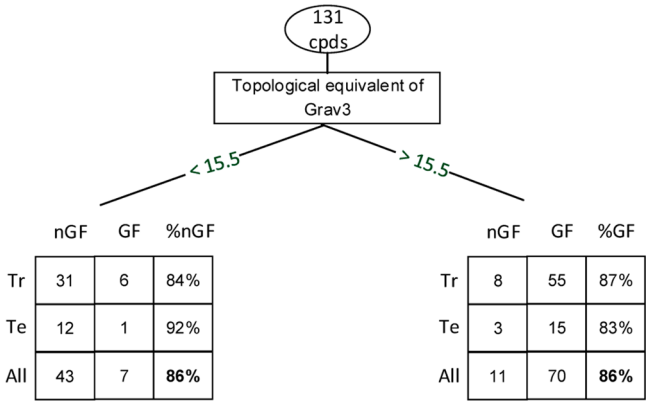


Figure 8. Prediction of the glass-forming ability using the support vector machine algorithm for all compounds under study: (a) training set and (b) test set. Abbreviations used: glass-former (GF) and nonglass-former (nGF).



	nGF	GF	%nGF
Tr	31	6	84%
Te	12	1	92%
All	43	7	**86%**

	nGF	GF	%GF
Tr	8	55	87%
Te	3	15	83%
All	11	70	**86%**

Figure 9. Prediction of the glass-forming ability of the studied compounds using the molecular descriptor "topological equivalent of Grav3" alone: (a) training set and (b) test set. Abbreviations used: compounds (cpds); glass-former (GF); nonglass-former (nGF); training set (Tr); and test set (Te).

300 g/mol are GFs. It has previously been shown that griseofulvin can be made amorphous by melt-cooling using a more rapid cooling rate of 40 °C/min in a DSC,[14] and so the resulting crystalline griseofulvin produced by the method employed herein conflicts with previous reports. A faster cooling than the methodology herein used could help in vitrifying the melt-quenched liquid. Hence, literature suggests that griseofulvin is correctly predicted by the MW cutoff for GFs, although we, in this study, were not able to produce the

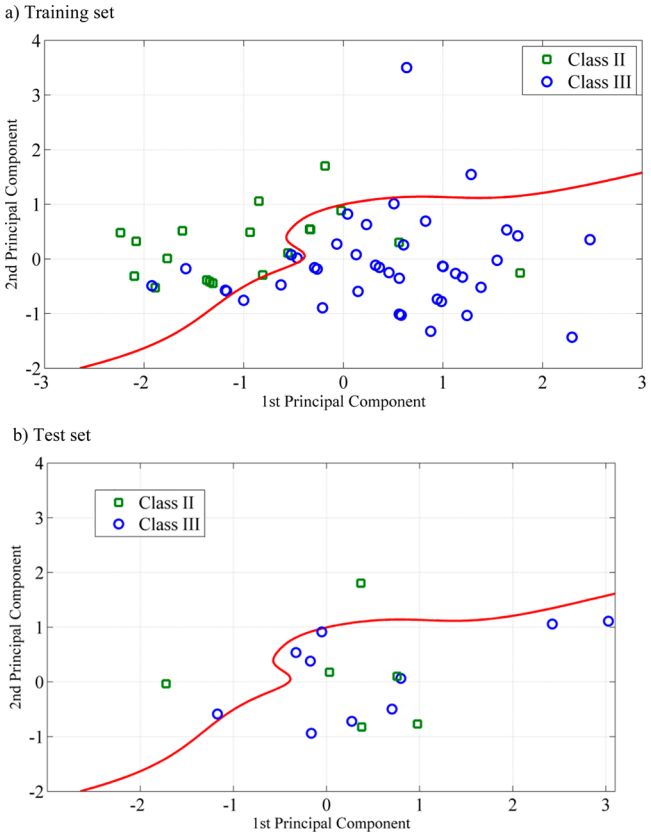


Figure 10. Prediction of the crystallization tendency of Class II and III compounds from calculated molecular descriptors.

amorphous material with the method applied. Our literature review did not indicate how to make amorphous forms of haloperidol, niclosamide, and spiperone. While we strongly believe that the cutoff values, decision trees, and SVMs reported herein will be very useful tools for the prediction of GFA, it needs to be emphasized that the method of preparation will also affect the GFA and crystallization properties of the amorphous materials.[22]

Prediction of GFA without MW Preference. In order to avoid the possible bias associated with the finding that MW is a strong predictor of the GFA, we iterated the data analysis in a second step using all the available descriptors. Two variables, reflecting the topological equivalent of Grav3_3D and the presence of ring structures (except for benzene and its condensed rings; i.e., aromatic, heteroaromatic, and hydrocarbon rings), predicted the GFA with an accuracy of 87% in both the training and test sets (Figure 8). Both these properties were related positively to GFA. Evaluation of the GFA using only the topological equivalent of Grav3_3D was also performed, and it was found that this descriptor alone differentiated the nGF compounds (84% accuracy) from the GFs (87%). Most importantly, strong predictions were also obtained for the test set (Figure 9). The GRAV-3 descriptor reflects the size, but also, shape and bulk properties of the molecule are captured by this descriptor.[23] The GRAV-3 is a cube root of gravitational index, and more details about calculating this descriptor can be found in Wessel et al.[23]

Prediction of Crystallization Tendency. The crystallization tendency of a compound is related to its solid-state behavior and stability in the solid state. Although the MW is a good predictor of Class I compounds with a MW less than 200

dx.doi.org/10.1021/mp500303a | *Mol. Pharmaceutics* 2014, 11, 3123−3132

g/mol, it was not possible to separate compounds from Classes II and III using MW. A model was therefore developed that would predict which molecules were in Classes II and III, to allow early identification of stable amorphous materials and amorphous states likely to need stabilizers. A model based on two descriptors, reflecting the sum of the absolute values of the Hückel pi atomic charges for only C atoms and the number of hydrogen bond acceptors, was developed. Figure 10 shows that Class II and Class III compounds were predicted in the training and test data with accuracies of 78% and 69%, respectively, using these descriptors. Hence, it appears that it is more difficult to predict the stability of the formed amorphous material than the GFA with computational tools. A high value for the sum of the absolute values of the Hückel pi atomic charges for only C atoms and the number of hydrogen bond acceptors predicts that the drug will be a Class III compound.[24] The Hückel method describes the delocalization of electrons. In addition, it has been suggested that hydrogen bonds are also an important requirement for compounds to form stable amorphous materials since the hydrogen bonds tend to increase the rigidity of the material, which otherwise has high mobility.[25]

CONCLUSIONS

In this study, the ability of 131 drugs to form glass and the crystallization tendency of the formed amorphous materials were investigated. SVMs were used to establish qualitative models distinguishing (i) GFs from nGFs and (ii) compounds that crystallize when the melt-quenched material is heated (Class II compounds) from compounds that do not crystallize under these conditions (Class III compounds). It was confirmed that a MW above 300 g/mol was an accurate indicator of GF compounds, and a MW of <200 g/mol was in this study was found to be a valid cutoff value for nGFs. However, MWs in the range of 200−300 g/mol cannot be used to predict GFA. The "topological equivalent of Grav3_3D" descriptor, which is related to the size and shape of the molecule, was found to be preferable to MW since it allowed prediction of GFA with 86% accuracy regardless of the MW. Of the 131 compounds investigated in this study, 77 compounds were found to be GFs. It was possible to predict the crystallization tendency of these drugs when in the amorphous state using two descriptors, reflecting the Hückel pi atomic charges and the number of hydrogen bond acceptors. To summarize, the results of the study confirmed the possibility of predicting the GFA of drug molecules with high accuracy using simple models. It was also shown for the first time that the crystallization tendency can be predicted without laboratory experiments. These findings provide tools for the rationalization of the current time- and material-consuming experimental setting, allowing the choice of possible formulation strategies to be guided by computational analysis of which poorly soluble drug molecules would be suitable as amorphous dosage forms.

AUTHOR INFORMATION

Corresponding Author

*E-mail: christel.bergstrom@farmaci.uu.se. Phone: +46 18 471 4118. Fax: +46 18 471 4223.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support for this project from the Swedish Research Council (Grants 621-2008-3777 and 621-2011-2445) is gratefully acknowledged. We would also like to thank Elisabeth and Alfred Ahlqvist for the postdoc grant to A.Alh. and VINNOVA for financial support for A.Alz. W.K. thanks Teva Pharmaceutical Industries LTD, USA, for additional financial support. We are also grateful for Simulations Plus (Lancaster, CA) providing the Drug Delivery Group at the Department of Pharmacy, Uppsala University, with a reference site license for the software ADMET Predictor.

REFERENCES

(1) Cui, Y. A material science perspective of pharmaceutical solids. *Int. J. Pharm.* **2007**, *339* (1), 3−18.
(2) Alhalaweh, A.; George, S.; Basavoju, S.; Childs, S. L.; Rizvi, S. A.; Velaga, S. P. Pharmaceutical cocrystals of nitrofurantoin: screening, characterization and crystal structure analysis. *CrystEngComm* **2012**, *14* (15), 5078−5088.
(3) Morissette, S. L.; Almarsson, Ö.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. High-throughput crystallization: polymorphs, salts, co-crystals and solvates of pharmaceutical solids. *Adv. Drug Delivery Rev.* **2004**, *56* (3), 275−300.
(4) Allesø, M.; Chieng, N.; Rehder, S.; Rantanen, J.; Rades, T.; Aaltonen, J. Enhanced dissolution rate and synchronized release of drugs in binary systems through formulation: Amorphous naproxen−cimetidine mixtures prepared by mechanical activation. *J. Controlled Release* **2009**, *136* (1), 45−53.
(5) Löbmann, K.; Laitinen, R.; Grohganz, H.; Gordon, K. C.; Strachan, C.; Rades, T. Coamorphous drug systems: enhanced physical stability and dissolution rate of indomethacin and naproxen. *Mol. Pharmaceutics* **2011**, *8* (5), 1919−1928.
(6) Aaltonen, J.; Allesø, M.; Mirza, S.; Koradia, V.; Gordon, K. C.; Rantanen, J. Solid form screening: a review. *Eur. J. Pharm. Biopharm.* **2009**, *71* (1), 23−37.
(7) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: an extraordinary example of conformational polymorphism. *Pharm. Res.* **2001**, *18* (6), 859−866.
(8) Byrn, S.; Pfeiffer, R.; Ganey, M.; Hoiberg, C.; Poochikian, G. Pharmaceutical solids: a strategic approach to regulatory considerations. *Pharm. Res.* **1995**, *12* (7), 945−954.
(9) Alhalaweh, A.; Kaialy, W.; Buckton, G.; Gill, H.; Nokhodchi, A.; Velaga, S. P. Theophylline cocrystals prepared by spray drying: physicochemical properties and aerosolization performance. *AAPS PharmSciTech* **2013**, *14* (1), 265−276.
(10) Yu, L. Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Adv. Drug Delivery Rev.* **2001**, *48* (1), 27−42.
(11) Hancock, B. C.; Parks, M. What is the true solubility advantage for amorphous pharmaceuticals? *Pharm. Res.* **2000**, *17* (4), 397−404.
(12) Baird, J. A.; Van Eerdenbrugh, B.; Taylor, L. S. A classification system to assess the crystallization tendency of organic molecules from undercooled melts. *J. Pharm. Sci.* **2010**, *99* (9), 3787−3806.
(13) Mahlin, D.; Bergström, C. A. Early drug development predictions of glass-forming ability and physical stability of drugs. *Eur. J. Pharm. Sci.* **2013**, *49* (2), 323−332.
(14) Mahlin, D.; Ponnambalam, S.; Heidarian Höckerfelt, M.; Bergström, C. A. Toward in silico prediction of glass-forming ability from molecular structure alone: a screening tool in early drug development. *Mol. Pharmaceutics* **2011**, *8* (2), 498−506.
(15) Nascimento, M. L.; Souza, L. A.; Ferreira, E. B.; Zanotto, E. D. Can glass stability parameters infer glass forming ability? *J. Non-Cryst. Solids.* **2005**, *351* (40), 3296−3308.
(16) Lu, Z.; Liu, C. A new glass-forming ability criterion for bulk metallic glasses. *Acta. Mater.* **2002**, *50* (13), 3501−3512.

(17) Avramov, I.; Zanotto, E.; Prado, M. Glass-forming ability versus stability of silicate glasses. II. Theoretical demonstration. *J. Non-Cryst. Solids.* **2003**, *320* (1), 9−20.
(18) Cabral, A., Jr.; Fredericci, C.; Zanotto, E. A test of the Hruby parameter to estimate glass-forming ability. *J. Non-Cryst. Solids.* **1997**, *219*, 182−186.
(19) Hearst, M. A.; Dumais, S.; Osman, E.; Platt, J.; Scholkopf, B. Support vector machines. *Intelli. Syst. Appl., IEEE* **1998**, *13* (4), 18−28.
(20) Fukui, K.; Yonezawa, T.; Shingu, H. A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Chem. Phys.* **1952**, *20* (4), 722−725.
(21) Li, T.; Ayers, P. W.; Liu, S.; Swadley, M. J.; Aubrey-Medendorp, C. Crystallization force: a density functional theory concept for revealing intermolecular interactions and molecular packing in organic crystals. *Chem. Eur. J.* **2009**, *15* (2), 361−371.
(22) Surana, R.; Pyne, A.; Suryanarayanan, R. Effect of preparation method on physical properties of amorphous trehalose. *Pharm. Res.* **2004**, *21* (7), 1167−1176.
(23) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of human intestinal absorption of drug compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (4), 726−735.
(24) Hückel, E. Quantentheoretische beiträge zum benzolproblem. *Z. Phys. Chem.* **1931**, *70* (3), 204−286.
(25) Kaushal, A. M.; Chakraborti, A. K.; Bansal, A. K. FTIR studies on differential intermolecular association in crystalline and amorphous states of structurally related non-steroidal anti-inflammatory drugs. *Mol. Pharmaceutics* **2008**, *5* (6), 937−945.