Original Article

Prediction of Cocrystal Formation Between Drug and Coformer by Simple Structural Parameters

Shadi Shayanfar[1], Abolghasem Jouyban[2], Sitaram Velaga[3], Ali Shayanfar[4,5]

[1]Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran, [2]Pharmaceutical Analysis Research Center, Tabriz University of Medical Sciences, Tabriz, Iran, [3]Pharmaceutical Engineering Group, Department of Health Sciences, Lulea University of Technology, Lulea, Sweden, [4]Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran, [5]Editorial Office of Pharmaceutical Sciences Journal, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract

Background: Cocrystal formation between an active pharmaceutical ingredient (API) and coformer is an applicable technique to change the physicochemical and pharmacokinetic properties. Computational methods can overcome the need for extensive experiments and improve the chances of success in the coformer selection. In this method, two compounds connect by non-covalent interactions that form a unique crystalline structure. Prediction of a cocrystal formation between API and coformer can help in the screening and design of new cocrystals. **Methods:** In this study, available data in the literature were applied to develop a prediction method based on binary logistic regression to screen cocrystal formation by sum and absolute difference of structural parameters (the number of rotatable bonds, Abraham solvation parameters, and topological polar surface area) of the two involved compounds. **Results:** The results showed various factors (eight structural parameters of the two compounds) could affect cocrystal formation, and the developed model can predict cocrystallization with a probability of about 90%. **Conclusion:** The related parameter to hydrogen bonding basicity and volume of compounds has the most significant effect on cocrystal formation.

Keywords: *Abraham solvation parameters, cocrystal, prediction*

Introduction

Physicochemical properties of active pharmaceutical ingredients (API), such as hygroscopicity, thermal stability, solubility, dissolution rate, and bioavailability, have an essential part in drug discovery and development.[1] Unfortunately, most of the candidate APIs have low solubility leading to poor bioavailability. To overcome this limitation, numerous methods have been recommended to increase solubility, such as cocrystal formation.[2] Cocrystallization technology has caught much attention in the pharmaceutical industry, and it can improve the properties of an API, such as solubility, bioavailability, storage stability, and manufacturability.[3,4] It allows better products to be produced for the marketplace by improving the physical and chemical properties of API without any change in structure, contrary to the situation that happened during salt formation.[5-7] The advantage of cocrystallization is that non-ionizable API molecules can form cocrystal against the salt formation, so the list of acceptable crystal coformers is more comprehensive than salt formation.[8]

Cocrystals are structurally homogeneous crystalline materials consisting of two molecules that are present in definite stoichiometric ratios, interacting with each other via a non-covalent bond, especially hydrogen bonding. The coformers are usually selected from Everything Added to Food in the United States (EAFUS) and the Generally Recognized As Safe (GRAS) lists. The fundamental rule of selecting coformers is based on the possibility of forming supramolecular synthons via intermolecular interactions.[9]

Based on the FDA guidance, cocrystals are a kind of solvates, whereas the second component is non-volatile. Some functional groups would form strong and directional hydrogen bonds, such as amides, heterocyclic bases, and carboxylic acids. These groups in the structures of APIs and coformers could lead to the formation of cocrystals.[1]

Experimental cocrystal preparation and characterization are time-consuming, and it needs relatively expensive instrumental

Received: 08 Dec 2021
Accepted: 18 Jun 2022
Published: 23 Dec 2022

Address for correspondence:
Dr. Ali Shayanfar, Editorial Office of Pharmaceutical Sciences Journal, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran.
E-mail: shayanfara@tbzmed.ac.ir

Access this article online

Website: www.jrpsjournal.com

DOI:10.4103/jrptps.JRPTPS_172_21

Quick Response Code:

How to cite this article: Shayanfar S, Jouyban A, Velaga S, Shayanfar A. Prediction of cocrystal formation between drug and coformer by simple structural parameters. J Rep Pharma Sci 2022;11:182-91.

methods such as powder X-ray diffraction and differential scanning calorimetry (DSC).[10]

Cocrystal screening and the possibility of cocrystal formation between API and coformer have become an increasingly important part of crystal engineering over the recent years. It can reduce required experiments to find out an appropriate coformer for an API.[11] There are many reports in the literature in which the attempts of researchers to prepare a cocrystal were unsuccessful and binary eutectic compositions obtained.[12-16] Investigation of factors affecting crystallization, computational approaches by various structural parameters for predicting cocrystal formation, and screening by different methods have been advanced in recent years to predict cocrystal formation. Structural parameters such as shape and polarity, Hansen solubility parameters of API[17] and coformer, *ab initio* molecular dynamics method,[18] conductor-like screening model for real solvents,[19,20] molecular electrostatic potential surfaces to assess molecular complementarity between two cocrystal components,[21] and realistic lattice energy landscapes[22] were suggested as essential parameters in cocrystal formation. In some cases, new statistical methods, e.g., multivariate adaptive regression splines methodology, were used for developing a prediction model.[23] In the majority of the *in silico* methods, a few compounds have been applied. Recently, descriptors based on fingerprint vectors and molecular graphs were applied for developing a model by artificial neural networks for a data set extracted from the Cambridge Structural Database.[24]

Valid and comprehensive experimental data are necessary to predict new cocrystal formation. One of the main reasons why this method is not mature and accurate to provide the acceptable prediction model is inaccessibility of comprehensive and uniform data sets (true positive/true negative cocrystal formation data). Przybyłek and Cysewski[25] have reported a data set comprising 226 binary mixtures (drug and phenolic acids as potential pharmaceutical coformers) classified as cocrystals or simple eutectics. They applied a multi-parameter model comprising 1D and 2D descriptors for cocrystal prediction with a probability of about 80%.

Simple statistical models to develop a mechanistic model, e.g., based on logistic regression and descriptors [i.e., topological polar surface area (TPSA), number of rotatable bonds (NRBs), and Abraham solvation parameters of components], are more acceptable in modeling studies[26] and could be useful in the future in cocrystal screening and formation. In this study, a virtual cocrystal screening process was reported on the basis of the structural parameters and logistic regression, and the prediction validity of the method was checked by leave-many-out cross-validation.

Computational Method

A cocrystal formation data set (true positive/true negative) comprising 226 data points was taken from a published article [Table 1].[25] SMILES (Simplified Molecular-Input Line-Entry System Codes) of compounds, a chemical notation that can be used by the computer that represents a chemical structure in a way,[27] were collected by searching their chemical names in PubChem (https://pubchem.ncbi.nlm.nih.gov). The numerical values for TPSA, NRB, and Abraham solvation parameters were computed by ACD/Labs software (https://ilab.acdlabs.com). Abraham solvation parameters are independent variables with the following solute properties: E is the excess molar refraction, S indicates dipolarity/polarizability descriptors of the solute, A is hydrogen-bond acidity, B is the solute hydrogen-bond acidity and basicity, and V is the McGowan volume of the solute.[28] Moreover, $A \times B$ was computed as another parameter which dealt with solute–solute interactions.[29] Then, sum and absolute difference of compounds 1 and 2's parameters were calculated.

The binary classification carried out using experimental true positive/true negative cocrystal formation was set as a dependent variable and sum and absolute difference of compounds 1 and 2's descriptors, i.e., TPSA, NRB, and Abraham solvation parameters, set as independent variables to develop a model based on logistic regression[30] to predict cocrystal formation. Logistic regression models compute the probabilities (P) for classification problems with two possible outcomes, i.e., $P>0.5$, class I (cocrystal formation) and $P<0.5$ (no cocrystal formation).

Statistically significant descriptors were selected for the logistic regression model based on probability values (P-value) associated with each descriptor whenever they were statistically significant at the 90% level ($P<0.1$). It displays the probability that the descriptor is by chance less than 10%.[26] The validity of the developed model was checked by the leave-10-out cross-validation method. In modeling studies, training a model by linear and non-linear models is not enough to confirm the prediction capability. The developed model should be applied to other data sets which did not include in the training of the model. On the way, whenever we can say a model is acceptable, it could predict the outcome of other compounds with reasonable accuracy. Therefore, external validation (test set) is essential. However, because of a small sample of available data, cross-validation was used. Cross-validation analysis, e.g., leave-many-out, is recommended in quantitative structure–activity studies, especially when the sample size is small, and some reports showed its superiority in external validation.[31,32] Therefore, a type of leave-many-out cross-validation (10 out) was applied. It evaluates the prediction capability of the model in which 10 compounds are left out from the training set, and the trained model was used to predict the removed data points. Figure 1 illustrated a schematic representation of

Table 1: Experimental cocrystal formation (true positive=1, true negative=0) between compounds 1 and 2[18] and probability (*P*) of a binary response by Eq. (1) (*P*>0.5: cocrystal formation, *P*<0.5: no cocrystal formation) and prediction cocrystal formation

No.	Experimental cocrystal formation	Compound 1	Compound 2	P	Prediction cocrystal formation
1	0	2,4,6-Trinitrotoluene	2-Hydroxybenzoic acid	0.077	0
2	0	2,4,6-Trinitrotoluene	Benzoic acid	0.024	0
3	0	2,4-Dinitrotoluene	2-Hydroxybenzoic acid	0.075	0
4	0	2,4-Dinitrotoluene	Benzoic acid	0.023	0
5	0	2-Chlorobenzoic acid	2-Hydroxybenzoic acid	0.249	0
6	0	2-Chlorobenzoic acid	Benzoic acid	0.089	0
7	0	2-Nitrophenol	Benzoic acid	0.406	0
8	0	3-Hydroxybenzamide	2-Hydroxybenzoic acid	0.695	1
9	0	3-Hydroxybenzamide	Benzoic acid	0.340	0
10	0	3-Methylheptanedioic acid	Benzoic acid	0.005	0
11	0	4-Nitrophenol	2-Hydroxybenzoic acid	0.420	0
12	0	4-Nitrophenol	Benzoic acid	0.180	0
13	0	4-Nitrophenol	Cinnamic acid	0.121	0
14	0	Acetamide	Benzoic acid	0.079	0
15	0	Benzamide	3-Hydroxybenzoic acid	0.644	1
16	0	Benzamide	4-Hydroxybenzoic acid	0.697	1
17	0	Benzamide	Benzoic acid	0.137	0
18	0	Benzamide	Ferulic acid	0.713	1
19	0	Benzyl	Benzoic acid	0.021	0
20	0	Benzoic acid	2-Hydroxybenzoic acid	0.242	0
21	0	Borneol	2-Hydroxybenzoic acid	0.343	0
22	0	Cinnamic acid	Benzoic acid	0.058	0
23	0	Curcumin	2-Hydroxybenzoic acid	0.271	0
24	0	Curcumin	4-Hydroxybenzoic acid	0.187	0
25	0	Curcumin	Ferulic acid	0.642	1
26	0	Ethenzamide	3-Hydroxybenzoic acid	0.567	1
27	0	Indomethacin	1-Hydroxy-2-naphthoic acid	0.929	1
28	0	Indomethacin	2,5-Dihydroxybenzoic acid	0.445	0
29	0	Indomethacin	2-Hydroxybenzoic acid	0.396	0
30	0	Indomethacin	4-Hydroxybenzoic acid	0.305	0
31	0	Indomethacin	Benzoic acid	0.185	0
32	0	Maleimide	2-Hydroxybenzoic acid	0.714	1
33	0	Maleimide	3-Hydroxybenzoic acid	0.931	1
34	0	Maleimide	Benzoic acid	0.498	0
35	0	Malonic acid	2-Hydroxybenzoic acid	0.329	0
36	0	Malonic acid	Benzoic acid	0.139	0
37	0	Menthol	2-Hydroxybenzoic acid	0.094	0
38	0	Naphthalene	Benzoic acid	0.015	0
39	0	Oxalic acid	Benzoic acid	0.192	0
40	0	Phenanthrene	2-Hydroxybenzoic acid	0.080	0
41	0	Phenanthrene	Benzoic acid	0.025	0
42	0	Phenanthrene	3-Hydroxybenzoic acid	0.321	0
43	0	Phenanthrene	Cinnamic acid	0.054	0
44	0	Phenylacetic acid	Benzoic acid	0.046	0
45	0	Phthalimide	Benzoic acid	0.544	1
46	0	Piracetam	2,6-Dihydroxybenzoic acid	0.946	1
47	0	Pregnenolone	2,5-Dihydroxybenzoic acid	0.185	0
48	0	Pregnenolone	2-Hydroxy-1-naphthoic acid	0.310	0
49	0	Pregnenolone	3-Hydroxy-2-naphthoic acid	0.310	0
50	0	Salicylamide	3-Hydroxybenzoic acid	0.807	1
51	0	Succinic acid	Benzoic acid	0.163	0
52	0	Urea	Benzoic acid	0.720	1

Table 1: Continued

No.	Experimental cocrystal formation	Compound 1	Compound 2	P	Prediction cocrystal formation
53	1	1,2-Diaminobenzene	Benzoic acid	0.878	1
54	1	1,4-Diaminobenzene	Benzoic acid	0.839	1
55	1	1h-Imidazole	4-Hydroxybenzoic acid	0.513	1
56	1	2-(4-(Dimethylamino)phenylazo) benzoic acid	2,6-Dihydroxybenzoic acid	0.984	1
57	1	2,3,5,6-Tetramethylpyrazine	4-Hydroxybenzoic acid	0.958	1
58	1	2,3,5,6-Tetramethylpyrazine	Ferulic acid	0.968	1
59	1	2,4,6-Trinitrophenol	Cinnamic acid	0.588	1
60	1	4-(1h-Pyrazol-1-ylmethyl) benzamide	Benzoic acid	0.134	0
61	1	4-(Pyridin-4-yl)pyridine 1-oxide	4-Hydroxycinnamic acid	0.981	1
62	1	4,4'-Bipyridine	2-Hydroxybenzoic acid	0.975	1
63	1	4,4'-Bipyridine	4-Hydroxybenzoic acid	0.996	1
64	1	4,4'-Bipyridine	6-Hydroxy-2-naphthoic acid	0.999	1
65	1	4-Amino-n-(4,6-dimethylpyrimidin)	2,4-Dihydroxybenzoic acid	0.995	1
66	1	4-Amino-n-(4,6-dimethylpyrimidin)	3-Hydroxy-2-naphthoic acid	1.000	1
67	1	4-Hydroxybenzamide	2-Hydroxybenzoic acid	0.586	1
68	1	4-Phenylpyridine	3-Hydroxybenzoic acid	0.819	1
69	1	4-Phenylpyridine	4-Hydroxybenzoic acid	0.853	1
70	1	Acetamiprid	4-Hydroxybenzoic acid	1.000	1
71	1	Acridine	3-Hydroxybenzoic acid	0.962	1
72	1	Acridine	4-Hydroxy-3-methoxybenzoic	0.995	1
73	1	Adenine	2-Hydroxybenzoic acid	0.973	1
74	1	Agomelatine	Benzoic acid	0.025	0
75	1	Benzamide	2,4-Dihydroxybenzoic acid	0.733	1
76	1	Benzamide	2,5-Dihydroxybenzoic acid	0.761	1
77	1	Benzamide	2,6-Dihydroxybenzoic acid	0.536	1
78	1	Benzamide	2-Hydroxybenzoic acid	0.297	0
79	1	Benzamide	3,5-Dihydroxybenzoic acid	0.838	1
80	1	Benzamide	3,4-Dihydroxybenzoic acid	0.903	1
81	1	Benzotriazole	4-Hydroxybenzoic acid	0.960	1
82	1	Caffeine	1-Hydroxy-2-naphthoic acid	1.000	1
83	1	Caffeine	2,3-Dihydroxybenzoic acid	1.000	1
84	1	Caffeine	2,4-Dihydroxybenzoic acid	1.000	1
85	1	Caffeine	2,5-Dihydroxybenzoic acid	1.000	1
86	1	Caffeine	2-Hydroxy-1-naphthoic acid	1.000	1
87	1	Caffeine	3,4,5-Trihydroxybenzoic acid	1.000	1
88	1	Caffeine	3,5-Dihydroxybenzoic acid	1.000	1
89	1	Caffeine	3-Hydroxy-2-naphthoic acid	1.000	1
90	1	Caffeine	4-Hydroxybenzoic acid	1.000	1
91	1	Caffeine	6-Hydroxy-2-naphthoic acid	1.000	1
92	1	Caffeine	Benzoic acid	1.000	1
93	1	Carbamazepine	2-Hydroxybenzoic acid	0.858	1
94	1	Carbamazepine	Cinnamic acid	0.789	1
95	1	Chlorzoxazone	2,4-Dihydroxybenzoic acid	0.996	1
96	1	Chlorzoxazone	4-Hydroxybenzoic acid	0.994	1
97	1	Danazol	4-Hydroxy-3-methoxybenzoic	0.689	1
98	1	Deferiprone	2,5-Dihydroxybenzoic acid	1.000	1
99	1	Deferiprone	4-Hydroxybenzoic acid	1.000	1
100	1	Didanosine	2-Hydroxybenzoic acid	1.000	1
101	1	Epoxiconazole	4-Hydroxybenzoic acid	0.790	1
102	1	Ethenzamide	2,4-Dihydroxybenzoic acid	0.724	1
103	1	Ethenzamide	2,5-Dihydroxybenzoic acid	0.754	1

Table 1: Continued

No.	Experimental cocrystal formation	Compound 1	Compound 2	P	Prediction cocrystal formation
104	1	Ethenzamide	2,6-Dihydroxybenzoic acid	0.526	1
105	1	Ethenzamide	2-Hydroxybenzoic acid	0.356	0
106	1	Ethenzamide	3,4-Dihydroxybenzoic acid	0.899	1
107	1	Ethenzamide	3,5-Dihydroxybenzoic acid	0.832	1
108	1	Ethenzamide	4-Hydroxy-3-methoxybenzoic	0.906	1
109	1	Ethenzamide	4-Hydroxybenzoic acid	0.626	1
110	1	Ethenzamide	Benzoic acid	0.161	0
111	1	Ethenzamide	Ferulic acid	0.893	1
112	1	Etodolac	2-Hydroxybenzoic acid	0.314	0
113	1	Etodolac	Ferulic acid	0.782	1
114	1	Etofylline	Benzoic acid	0.999	1
115	1	Ferulic acid	2-Hydroxybenzoic acid	0.669	1
116	1	Fluconazole	2-Hydroxybenzoic acid	0.952	1
117	1	Flucytosine	Benzoic acid	1.000	1
118	1	Gabapentin	1-Hydroxy-2-naphthoic acid	0.981	1
119	1	Gabapentin	2-Hydroxybenzoic acid	0.918	1
120	1	Gabapentin	3-Hydroxybenzoic acid	0.741	1
121	1	Gabapentin	4-Hydroxybenzoic acid	0.790	1
122	1	Iloperidone	2,3-Dihydroxybenzoic acid	0.972	1
123	1	Iloperidone	3,5-Dihydroxybenzoic acid	0.988	1
124	1	Iloperidone	3-Hydroxybenzoic acid	0.958	1
125	1	Imazamox	4-Hydroxybenzoic acid	1.000	1
126	1	Imazethapyr	4-Hydroxybenzoic acid	1.000	1
127	1	Isoniazid	3,4,5-Trihydroxybenzoic acid	1.000	1
128	1	Isoniazid	3,5-Dihydroxybenzoic acid	1.000	1
129	1	Isoniazid	3-Hydroxybenzoic acid	0.999	1
130	1	Isoniazid	4-Hydroxy-3-methoxybenzoic	1.000	1
131	1	Isoniazid	Caffeic acid	1.000	1
132	1	Isoniazid	Ferulic acid	1.000	1
133	1	Isoniazid	2,3-Dihydroxybenzoic acid	1.000	1
134	1	Isonicotinamide	2-Hydroxybenzoic acid	0.909	1
135	1	Isonicotinamide	Benzoic acid	0.821	1
136	1	Isonicotinonitrile	2,5-Dihydroxybenzoic acid	0.906	1
137	1	Lornoxicam	2-Hydroxybenzoic acid	1.000	1
138	1	Maleimide	2,4-Dihydroxybenzoic acid	0.954	1
139	1	Maleimide	3,4,5-Trihydroxybenzoic acid	0.981	1
140	1	Maleimide	3,4-Dihydroxybenzoic acid	0.983	1
141	1	Maleimide	3,5-Dihydroxybenzoic acid	0.967	1
142	1	Maleimide	4-Hydroxybenzoic acid	0.945	1
143	1	Meloxicam	1-Hydroxy-2-naphthoic acid	1.000	1
144	1	Meloxicam	2,5-Dihydroxybenzoic acid	1.000	1
145	1	Meloxicam	2-Hydroxybenzoic acid	1.000	1
146	1	Meloxicam	4-Hydroxybenzoic acid	1.000	1
147	1	Meloxicam	Benzoic acid	1.000	1
148	1	Metronidazole	3,4,5-Trihydroxybenzoic acid	0.998	1
149	1	n,n′-Diacetylpiperazine	2-Hydroxybenzoic acid	0.984	1
150	1	Nevirapine	3,4,5-Trihydroxybenzoic acid	1.000	1
151	1	Nevirapine	4-Hydroxybenzoic acid	0.998	1
152	1	Nevirapine	Benzoic acid	0.992	1
153	1	Nevirapine	Ferulic acid	1.000	1
154	1	Nicotinamide	2,5-Dihydroxybenzoic acid	0.973	1
155	1	Nicotinamide	2-Hydroxybenzoic acid	0.909	1
156	1	Nicotinamide	3-Hydroxy-2-naphthoic acid	0.972	1

Table 1: Continued

No.	Experimental cocrystal formation	Compound 1	Compound 2	P	Prediction cocrystal formation
157	1	Nicotinamide	3-Hydroxybenzoic acid	0.953	1
158	1	Nicotinamide	4-Hydroxybenzoic acid	0.963	1
159	1	Nicotinamide	4-Hydroxycinnamic acid	0.966	1
160	1	Nicotinamide	Sinapic acid	0.982	1
161	1	Nicotinamide	2,6-Dihydroxybenzoic acid	0.928	1
162	1	Nicotinamide	3,4-Dihydroxybenzoic acid	0.990	1
163	1	Nicotinamide	3,5-Dihydroxybenzoic acid	0.983	1
164	1	Nicotinamide	Ferulic acid	0.965	1
165	1	Paliperidone	4-Hydroxybenzoic acid	1.000	1
166	1	Paliperidone	4-Hydroxybenzoic acid	1.000	1
167	1	Pentoxifylline	3,4,5-Trihydroxybenzoic acid	1.000	1
168	1	Phenazine	2,6-Dihydroxybenzoic acid	0.995	1
169	1	Piperazine	2-Hydroxybenzoic acid	0.985	1
170	1	Piperazine-2,5-dione	3,5-Dihydroxybenzoic acid	1.000	1
171	1	Piracetam	2,5-Dihydroxybenzoic acid	0.992	1
172	1	Piracetam	3-Hydroxybenzoic acid	0.982	1
173	1	Piracetam	4-Hydroxybenzoic acid	0.986	1
174	1	Piroxicam	2,5-Dihydroxybenzoic acid	1.000	1
175	1	Piroxicam	2-Hydroxybenzoic acid	1.000	1
176	1	Piroxicam	4-Hydroxybenzoic acid	1.000	1
177	1	Piroxicam	Enzoic acid	1.000	1
178	1	Progesterone	2-Hydroxy-1-naphthoic acid	0.405	0
179	1	Prulifloxacin	2-Hydroxybenzoic acid	1.000	1
180	1	Pyraclostrobin	4-Hydroxybenzoic acid	0.981	1
181	1	Pyrazinamide	1-Hydroxy-2-naphthoic acid	0.996	1
182	1	Pyrazinamide	3,4-Dihydroxybenzoic acid	0.998	1
183	1	Pyrazinamide	3-Hydroxybenzoic acid	0.992	1
184	1	Pyrazinamide	4-Hydroxy-3-methoxybenzoi	0.997	1
185	1	Pyrazinamide	4-Hydroxybenzoic acid	0.994	1
186	1	Salicylamide	2,4-Dihydroxybenzoic acid	0.885	1
187	1	Salicylamide	2,5-Dihydroxybenzoic acid	0.900	1
188	1	Salicylamide	2,6-Dihydroxybenzoic acid	0.765	1
189	1	Salicylamide	3,4-Dihydroxybenzoic acid	0.963	1
190	1	Salicylamide	3,5-Dihydroxybenzoic acid	0.936	1
191	1	Salicylamide	4-Hydroxybenzoic acid	0.842	1
192	1	Salicylamide	Benzoic acid	0.280	0
193	1	Salicylamide	Ferulic acid	0.875	1
194	1	Stanozolol	6-Hydroxy-2-naphthoic acid	0.332	0
195	1	Tadalafil	2-Hydroxybenzoic acid	1.000	1
196	1	Theophylline	1-Hydroxy-2-naphthoic acid	1.000	1
197	1	Theophylline	2-Hydroxy-1-naphthoic acid	1.000	1
198	1	Theophylline	2-Hydroxybenzoic acid	0.999	1
199	1	Theophylline	3-Hydroxy-2-naphthoic acid	1.000	1
200	1	Trimethoprim	Benzoic acid	0.854	1
201	1	Urea	2,4-Dihydroxybenzoic acid	0.927	1
202	1	Urea	2,5-Dihydroxybenzoic acid	0.925	1
203	1	Urea	2,6-Dihydroxybenzoic acid	0.885	1
204	1	Urea	2-Hydroxybenzoic acid	0.896	1
205	1	Urea	3,5-Dihydroxybenzoic acid	0.913	1
206	1	Urea	3-Hydroxybenzoic acid	0.897	1
207	1	Urea	4-Hydroxybenzoic acid	0.918	1
208	1	Urea	3,4-Dihydroxybenzoic acid	0.954	1
209	1	Urea	Ferulic acid	0.847	1
210	1	Voriconazole	4-Hydroxybenzoic acid	0.994	1

Table 2: Percentage of correct prediction of cocrystal formation between compounds 1 and 2 by Eq. (1) and leave-10-out cross-validation

Cocrystal formation	Eq. (1)	Leave-10-out cross-validation
True positive	94% (149 correct prediction from 158 total data)	92% (145 correct prediction from 158 total data)
True negative	75% (39 correct prediction from 52 total data)	73% (38 correct prediction from 52 total data)
Overall	90% (188 data from 210 total data)	87% (183 correct prediction from 210 total data)

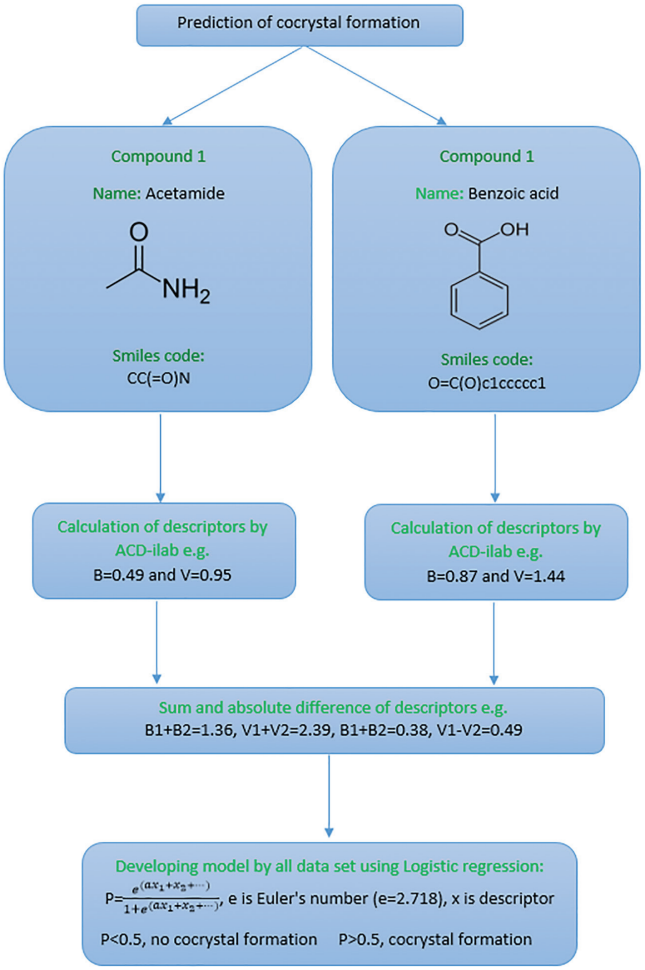

Figure 1: A schematic representation of computational methods for the calculation of parameters and logistic regression for predicting cocrystal formation in this study

computational methods for the calculation of parameters and logistic regression for predicting cocrystal formation in this study.

Results and Discussion

Based on the preliminary analysis, reported data of cocrystal formation between carboximide and carboxylic acid behave as outliers (17 data). Geometric disposition of hydroxyl functionality on benzoic acid that can drive the formation of cocrystals[33] is a possible reason to act as outliers, and structural parameters cannot estimate cocrystal formation. Therefore, remained data points (210 data) were applied to develop a model by sum and absolute difference of compounds 1 and 2's descriptors, i.e., TPSA, NRB, and Abraham solvation parameters based, on binary logistic regression, in which the obtained model is:

$$P=\frac{e^{-7.984-2.375|V_1-V_2|+0.068|TPSA_1-TPSA_2|-0.818|NRB_1+NRB_2|+13.236|B_1+B_2|-5.458|E_1+E_2|+3.855|V_1+V_2|-6.104|AB_1+AB_2|-2.287|S_1-S_2|-1.628|A_1-A_2|+2.740|AB_1-AB_2|}}{1+e^{-7.984-2.375|V_1-V_2|+0.068|TPSA_1-TPSA_2|-0.818|NRB_1+NRB_2|+13.236|B_1+B_2|-5.458|E_1+E_2|+3.855|V_1+V_2|-6.104|AB_1+AB_2|-2.287|S_1-S_2|-1.628|A_1-A_2|+2.740|AB_1-AB_2|}} \quad (1)$$

where P is the probability of binary responses (class 0: P<0.5 or 1 P>0.5) based on the true positive/true negative cocrystal formation and e is Euler's number (e=2.718). In addition, probability values (P-value) associated with each descriptor were less than 0.1. The model [Eq. (1)] estimated 94% and 75% of true positive and true negative cocrystal formation in the correct group, respectively [Table 1]. Leave-many-out cross-validation was performed to evaluate the prediction capability of the model, and the results were listed in Table 2. No significant difference in accuracy after leave-10-out cross-validation confirms the prediction capability of the developed model. Besides, receiver-operating characteristic (ROC) curves were used to evaluate the classification abilities of the logistic-regression-based models. ROC of Eq. (1) was illustrated in Figure 2, and the area under the curve is 0.943. It confirms the validity of the developed model for predicting cocrystal formation between the two compounds.

Evaluation of the selected parameters shows that the hydrogen bonding parameters have the most significant effect on cocrystal formation. Feature selection with a forward method using SPSS software indicated that the most crucial parameter for the classification of data is the sum of the hydrogen-bond basicity (B_1+B_2) and volume of compounds (V_1-V_2). With these two parameters [Eq. (2)], 60% and 95% of true positive and true negative cocrystal formations are classified in the correct group, respectively:

$$P=\frac{e^{-4.845+5.629|B_1+B_2|-0.450|V_1-V_2|}}{1+e^{-4.845+5.629|B_1+B_2|-0.450|V_1-V_2|}} \quad (2)$$

Based on the developed model, high numerical values correspond to hydrogen bond donor of drug and coformer and similar volume between the two compounds gives a P higher than 0.5, which indicates cocrystal formation. Further analyses showed that all of the true positive cocrystal formation data between compounds 1 and 2 have B_1+B_2>1 [Figure 3]. In contrast, this value for No cocrystal formation data is 66% and only three data have

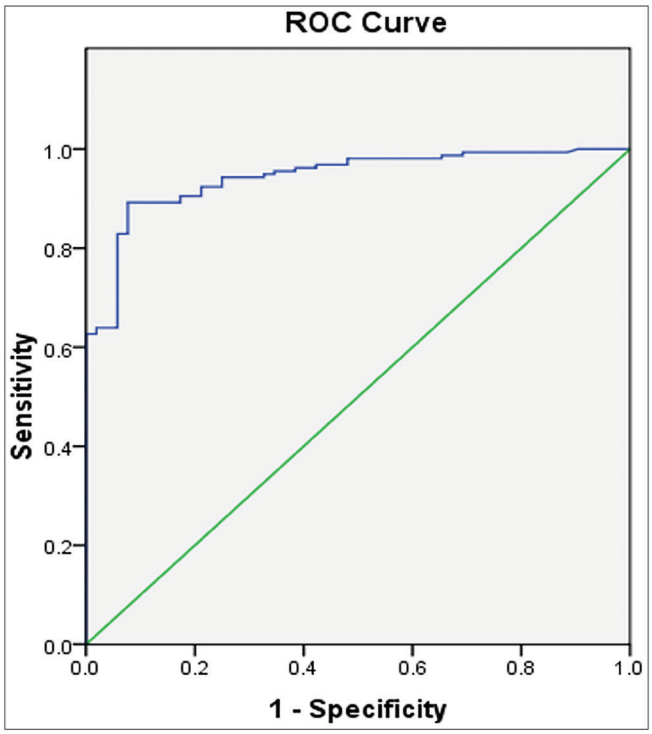

Figure 2: Receiver-operating characteristic (ROC) curve of the model (Eq. 1) for predicting cocrystal formation

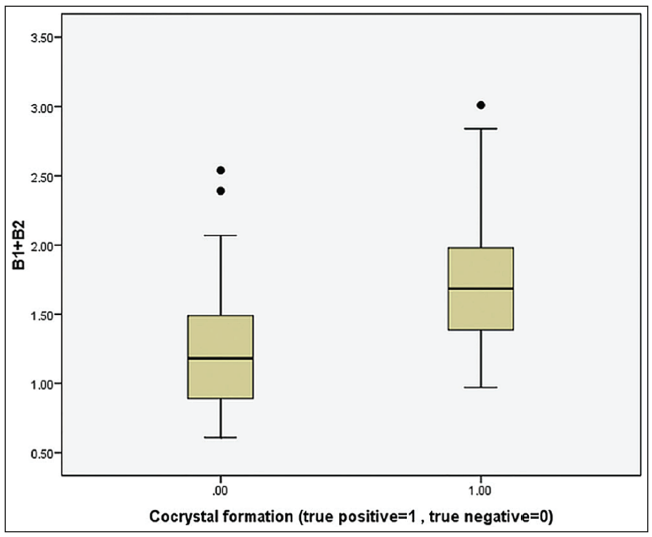

Figure 3: Cocrystal formation vs. B_1+B_2 of compounds 1 and 2. All of the true positive cocrystal formation between compounds 1 and 2 has $B_1+B_2>1$, whereas this value for no cocrystal formation data is 66% and only three data have $B_1+B_2>2$

$B_1+B_2>2$. The cocrystal can be formed using non-covalent interaction, which is often hydrogen bond.[34,35] It has a critical effect on the cocrystal formation between the two compounds. Typical hydrogen bonds utilized in crystal engineering are bonds among carboxylic acids, amides, and N-heterocyclic hydrogen-bond acceptor with carboxylic acid.[36] According to the studied parameters, *A* and *B* belong to the hydrogen bonding group of compounds, and they represent hydrogen bonding donor and acceptor

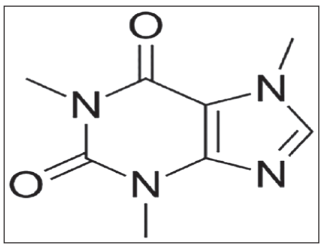

Figure 4: Structure of caffeine. It has no hydrogen bond donor functional group (*A*=0); however, they could form cocrystal with various coformers

groups, respectively. These results indicate that although "*A*" could have a proper role in cocrystal formation based on the obtained model [Eq. (1)], "*B*" in comparison with "*A*" is a more critical parameter affecting cocrystal formation. Some compounds with $A=0$ such as caffeine [Figure 4] which do not have hydrogen bond donor groups can form cocrystal with some of the coformers. Moreover, these results confirm the previous findings, in which supramolecular heterosynthons (i.e., non-covalent bonds between different functional groups composed of a hydrogen bond acceptor and a hydrogen bond donor group such as carboxylic acid–aromatic nitrogen) favor for cocrystal formation than supramolecular homosynthon.[37] Although the developed model [(Eq. (2)] with only two parameters predicts cocrystal formation with 86% accuracy, nevertheless, the prediction for No cocrystal formation is not acceptable (60%) because it is only 10% close to the chance probability value (50%). Other parameters were used to improve the accuracy values of true positive and true negative cocrystal formation to 94% and 75%, respectively [(Eq. (1)]. Based on the mentioned descriptors, the model did not have satisfying results unless the sum and absolute difference of another descriptor (sum and absolute difference of $A\times B$) dealt with acidity and basicity hydrogen bond descriptors of compounds included in the model yielded much better correlations. The obtained results verify the importance of hydrogen bonding in cocrystal formation. Similar outcomes to including $A\times B$ in developing quantitative structure–property relationship (QSPR) models have been reported by Abraham and Acree[29] for the solubility prediction of solutes in octanol and water by Abraham solvation parameters.

The results of this study approve the capability of Abraham solvation parameters for predicting cocrystal formation. These parameters have been extensively used in QSPR models to predict physicochemical and pharmacokinetic properties.[38-41] Moreover, the established models based on logistic regression using Abraham solvation parameters as structural descriptors have been applied to estimate Biopharmaceutics Drug Disposition Classification System (BDDCS)[42] and the oral bioavailability correlation between humans and preclinical animals.[43]

In this study, for the first time, Abraham solvation parameters and some other simple descriptors were used

for the prediction of non-covalent interaction between API and coformer, i.e., cocrystal formation. The results indicate the importance of related parameters to hydrogen binding basicity and volume of API and conformer, which is the most common interaction in cocrystal formation. Mechanistic interpretation for the developed model means that it is potentially helpful in the regulatory context.[26]

The limitation of this study is a relatively small data set (only 52 NO cocrystal formation data). The applied descriptors and the developed model on a large uniform data set composed of true positive and true negative cocrystal formation could improve the prediction accuracy of the model. Other structural descriptors, i.e., intermolecular descriptors and 1D, 2D, and even 3D descriptors, could have more positive effects on the accuracy of cocrystal formation prediction between drug and coformer. However, the developed models with them are complex, and mechanistic interpretation sometimes is not possible.

Conclusion

Prediction of cocrystal formation of a drug with different coformers helps to design the cocrystal in drug development, and it could speed up the process of cocrystal formation studies. Hydrogen bonding parameters and the molecules' volume are the most important in cocrystal formation between the two compounds. The molecules with similar volumes and more functional groups involved in hydrogen bonding are positive structural chemical factors in cocrystal formation. However, other parameters are necessary to propose an applicable model with good accuracy to predict cocrystal formation. The results of this study can help the mechanistic evaluation of cocrystal formation and cocrystal screening without time-consuming and costly experiment methods with only computational structural descriptors. However, finalization of cocrystal formation requires other standard methods such as DSC, X-ray diffractions, and so on.

Financial support and sponsorship

The authors would like to thank the Tabriz University of Medical Sciences for the financial support (60049) of the project.

Conflicts of interest

There are no conflicts of interest.

Authors' contribution

SS and AS performed data analysis and manuscript writing. SV contributed in data analysis and AJ contributed in manuscript preparation. All authors read and approved the final manuscript.

Ethics approval

Not applicable.

References

1. Du Y, Xue J, Cai Q, Zhang Q. Spectroscopic investigation on structure and pH dependent cocrystal formation between gamma-aminobutyric acid and benzoic acid. Spectrochim Acta A Mol Biomol Spectrosc 2018;191:377-81.
2. Emami S, Siahi-Shadbad M, Adibkia K, Barzegar-Jalali M. Recent advances in improving oral drug bioavailability by cocrystals. Bioimpacts 2018;8:305-20.
3. Kavanagh ON, Croker DM, Walker GM, Zaworotko MJ. Pharmaceutical cocrystals: From serendipity to design to application. Drug Discov Today 2019;24:796-804.
4. Dutt B, Choudhary M, Budhwar V. Cocrystallization: An innovative route toward better medication. J Rep Pharm Sci 2020;9:256-70.
5. Diniz LF, Souza MS, Carvalho PS Jr, da Silva CCP, D'Vries RF, Ellena J. Novel isoniazid cocrystals with aromatic carboxylic acids: Crystal engineering, spectroscopy and thermochemical investigations. J Mol Struct 2018;1153:58-68.
6. Alvani A, Jouyban A, Shayanfar A. The effect of surfactant and polymer on solution stability and solubility of tadalafil–methylparaben cocrystal. J Mol Liq 2019;281:86-92.
7. Lara-Ochoa F, Espinosa-Pérez G. Cocrystals definitions. Supramol Chem 2007;19:553-7.
8. Karimi-Jafari M, Padrela L, Walker GM, Croker DM. Creating cocrystals: A review of pharmaceutical cocrystal preparation routes and applications. Cryst Growth Des 2018;18:6370-87.
9. Manin AN, Drozd KV, Churakov AV, Perlovich GL. Hydrogen bond donor/acceptor ratios of the coformers: Do they really matter for the prediction of molecular packing in cocrystals? The case of benzamide derivatives with dicarboxylic acids. Cryst Growth Des 2018;18:5254-69.
10. Malamatari M, Ross SA, Douroumis D, Velaga SP. Experimental cocrystal screening and solution based scale-up cocrystallization methods. Adv Drug Deliv Rev 2017;117:162-77.
11. Kumar A, Nanda A. *In-silico* methods of cocrystal screening: A review on tools for rational design of pharmaceutical cocrystals. J Drug Deliv Sci Technol 2021;63:102527.
12. Goud NR, Suresh K, Sanphui P, Nangia A. Fast dissolving eutectic compositions of curcumin. Int J Pharm 2012;439:63-72.
13. Shayanfar A, Jouyban A. Physicochemical characterization of a new cocrystal of ketoconazole. Powder Technol 2014;262:242-8.
14. Bala M, Gautam MK, Chadha R. What if cocrystallization fails for neutral molecules? Screening offered eutectics as alternate pharmaceutical materials: Leflunomide—A case. Pharm Sci 2019;25:235-43.
15. Emami S, Siahi-Shadbad M, Barzegar-Jalali M, Adibkia K. Characterizing eutectic mixtures of gliclazide with succinic acid prepared by electrospray deposition and liquid assisted grinding methods. J Drug Deliv Sci Technol 2018;45:101-9.
16. Fandiño OE, Bruno FP, Monti GA, Sperandeo NR. Mechanochemical synthesis of a novel eutectic of the antimicrobial nitazoxanide with improved dissolution performance. Pharm Sci 2021;27:585-92.
17. Mohammad MA, Alhalaweh A, Velaga SP. Hansen solubility parameter as a tool to predict cocrystal formation. Int J Pharm 2011;407:63-71.
18. Fábián L. Cambridge structural database analysis of molecular complementarity in cocrystals. Cryst Growth Des 2009;9:1436-43.

19. Przybyłek M, Ziółkowska D, Mroczyńska K, Cysewski P. Applicability of phenolic acids as effective enhancers of cocrystal solubility of methylxanthines. Cryst Growth Des 2017;17: 2186-93.
20. Loschen C, Klamt A. Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. J Pharm Pharmacol 2015;67:803-11.
21. Grecu T, Prohens R, McCabe JF, Carrington EJ, Wright JS, Brammer L, *et al*. Cocrystals of spironolactone and griseofulvin based on an *in silico* screening method. CrystEngComm 2017;19:3592-9.
22. Mohamed S, Tocher DA, Price SL. Computational prediction of salt and cocrystal structures—Does a proton position matter? Int J Pharm 2011;418:187-98.
23. Przybyłek M, Jeliński T, Słabuszewska J, Ziółkowska D, Mroczyńska K, Cysewski P. Application of multivariate adaptive regression splines (MARSplines) methodology for screening of dicarboxylic acid cocrystal using 1D and 2D molecular descriptors. Cryst Growth Des 2019;19:3876-87.
24. Devogelaer JJ, Meekes H, Tinnemans P, Vlieg E, de Gelder R. Co-crystal prediction by artificial neural networks. Angew Chem Int Ed 2020;59:21711-8.
25. Przybyłek M, Cysewski P. Distinguishing cocrystals from simple eutectic mixtures: Phenolic acids as potential pharmaceutical coformers. Cryst Growth Des 2018;18:3524-34.
26. Dearden JC, Cronin MT, Kaiser KL. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Environ Res 2009;20:241-66.
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. Acree WE, Grubbs LM, Abraham MH. Prediction of Partition Coefficients and Permeability of Drug Molecules in Biological Systems with Abraham Model Solute Descriptors Derived From Measured Solubilities and Water-to-Organic Solvent Partition Coefficients, Toxicity and Drug Testing. Prediction of Partition Coefficients and Permeability of Drug Molecules in Biological Systems with Abraham Model Solute Descriptors, Derived from Measured Solubilities and Water-to-organic Solvent Partition Coefficients. London: Intech Publisher; 2012. p. 100-2.
29. Abraham MH, Acree WE Jr. The solubility of liquid and solid compounds in dry octan-1-ol. Chemosphere 2014;103:26-34.
30. Shimono E, Inoue K, Kurita T, Ichiraku Y. Logistic regression analysis for the material design of chiral crystals. Chem Lett 2018;47:611-2.
31. Majumdar S, Basak SC. Beware of external validation! A comparative study of several validation techniques used in QSAR modelling. Curr Comput-Aided Drug Des 2018;14:284-91.
32. Gütlein M, Helma C, Karwath A, Kramer S. A large-scale empirical evaluation of cross-validation and external test set validation in (Q)SAR. Mol Inform 2013;32:516-28.
33. Kaur R, Gautam R, Cherukuvada S, Guru Row TN. Do carboximide-carboxylic acid combinations form co-crystals? The role of hydroxyl substitution on the formation of co-crystals and eutectics. IUCRJ 2015;2:341-51.
34. Tothadi S, Shaikh TR, Gupta S, Dandela R, Vinod CP, Nangia AK. Can we identify the salt-cocrystal continuum state using XPS? Cryst Growth Des 2021;21:735-47.
35. Kimball D, Starnes J, Groeneman RH, Krueger HR, Reinheimer EW. Two-dimensional sheet based on C-H···N hydrogen bonds within an organic cocrystal: A crystallographic, spectroscopic and theoretical study. Supramol Chem 2015;27:465-70.
36. Williams HD, Trevaskis NL, Charman SA, Shanker RM, Charman WN, Pouton CW, *et al*. Strategies to address low drug solubility in discovery and development. Pharmacol Rev 2013;65:315-499.
37. Shan N, Zaworotko MJ. The role of cocrystals in pharmaceutical science. Drug Discov Today 2008;13:440-6.
38. Hoover KR, Acree WE Jr, Abraham MH. Chemical toxicity correlations for several fish species based on the Abraham solvation parameter model. Chem Res Toxicol 2005;18:1497-505.
39. Liu G, Eddula S, Jiang C, Huang J, Tirumala P, Xu A, *et al*. Abraham solvation parameter model: Prediction of enthalpies of vaporization and sublimation of mono-methyl branched alkanes using measured gas chromatographic data. Eur Chem Bul 2020;9:273-84.
40. Abraham MH, Austin RP. The effect of ionized species on microsomal binding. Eur J Med Chem 2012;47:202-5.
41. Jouyban A, Rahimpour E, Acree WE Jr. Further analysis on solubility measurement and thermodynamic modeling of benzoic acid in monosolvents and binary mixtures. Pharm Sci 2019;25:165-70.
42. Golfar Y, Shayanfar A. Prediction of biopharmaceutical drug disposition classification system (BDDCS) by structural parameters. J Pharm Pharm Sci 2019;22:247-69.
43. Beheshti S, Shayanfar A. Prediction of the oral bioavailability correlation between humans and preclinical animals. Eur J Drug Metab Pharmacokinet 2020;45:771-83.