

Д.И. Осолодкин: Вы упоминаете, что в обучающей выборке есть данные о растворителе, при этом нигде более в вашем докладе растворитель не фигурирует. Для какого растворителя построены ваши модели и учитываются ли каким-либо образом при их построении данные для других растворителей? Если да, то как именно?

А.Б. Рахимбекова: В наборах данных по константам таутомерного равновесия и кислотности измерения проведены в 24 и 13 различных растворителях, соответственно. Вектор дескрипторов для каждой объекта (КГР реакции или молекулы) был получен путем конкатенации вектора фрагментных дескрипторов объекта и дескрипторов, описывающих свойства растворителя и дескриптора температуры. В качестве дескрипторов растворителей использовали: константы Каталана: SPP (полярность/поляризуемость), SA (кислотность), SB (основность); константы Камлета-Тафта: α (донорная способность), β (акцепторная способность), π (поляризуемость); функции диэлектрической проницаемости растворителя (4 дескриптора) и показателя преломления растворителей (3 дескриптора). Для моделирования водно-органических смесей вводили также дескриптор, характеризующий мольное процентное содержание органического растворителя в смеси (для чистых органических растворителей равен 100). Дескриптором температуры служила обратная температура, выраженная в градусах Кельвина.

Д.И. Осолодкин: Насколько соответствуют друг другу обучающие выборки? По идее, данные по константам кислотности доступны для значительно большего набора соединений, чем данные по константам таутомерного равновесия, просто потому, что многие соединения не обладают таутомерией. Учитываются ли как-то эти особенности?

А.Б. Рахимбекова: Пересечений немного. Мы нашли 19 примеров, когда одно и то же соединение фигурировало в базе данных по кислотности и по таутомерии. Согласно публикациям, около 25% химических соединений таутомеризуемы. Но при измерении кислотности этим фактором обычно пренебрегают. Данных по кислотности действительно много, но в большинстве своем, кислотный центр в них не выделен, а известна только кислотность соединения (то есть макроконстанта кислотности). Это не подходит для нашей модели, нам нужны были данные, в которых кислотный центр известен (то есть данные по микроконстантам). Кроме того, данные по кислотности относятся к одному растворителю - воде. В то время как данные по таутомерии, собранные нами, относятся к разным растворителям. По этой причине мы сами собирали данные по кислотностям, а не использовали литературные. Для проверки нашего подхода сопряженных QSPR моделей этого объема данных вполне достаточно. Создание модели на основе всех имеющихся данных является целью дальнейшей работы