

# In Silico Techniques in Lead Optimization

## Perspectives for Rational Use

Pranas Japertas, Rytis Kubilius,  
Andrius Sazonovas, Kiril Lanevskij

ACD/Labs, Inc., A. Mickeviciaus g. 29,  
LT-08117 Vilnius, Lithuania



### INTRODUCTION

The efforts of lead optimization projects are directed towards analogs that have favorable ADME profiles and are devoid of safety concerns whilst retaining target activity. In this work we present a novel computational platform to aid such projects by generating virtual analog libraries, in the physicochemical space regions compatible with desired biological characteristics.

The main idea behind our approach is that many considered properties are governed by basic physicochemical parameters, such as ionization, lipophilicity, or molecular size. We have devised simple, yet accurate physicochemical models of intestinal absorption and passive permeation across the BBB, as well as general physicochemical rules that hold even for certain protein-ligand interactions. Changing parameter values may have distinct, even opposite effects on different ADME properties. The impact of a particular parameter may depend on the allowed variation ranges of other parameters. Using the cumulative output of available predictive models enables us to account for the multitude of possible effects, and identify the regions in physicochemical space most likely occupied by analogs with the desired combination of ADME properties. Advanced techniques are also applied to improve selection of substituents fitting within these regions, including custom Hammett equations for estimating the mutual effects of the core molecule and the modified substituent on the analog's  $pK_a$  values.

### PHYSICOCHEMICAL RULES FOR ADME PROPERTIES

All lead optimization project starts with definition of the desired ADME profile that the candidate compound must fulfill in order to become a viable drug. Many QSAR models are available to aid selection of potentially good candidates by producing estimates of the properties of interest directly from chemical structure. The majority of these models, however, are statistics-driven and function as 'black boxes' without any feedback about why a particular value was obtained.

Yet, many biological properties are governed by the simplest physicochemical characteristics, and different values can be attributed to certain combinations of e.g.  $\log P$  and  $pK_a$ . We have shown that passive permeability of chemicals across gastrointestinal and blood-brain barriers can be accurately predicted using only basic physicochemical parameters as inputs.<sup>1-4</sup>

Figs. 1-2 provide an example of this relationship. Here, the extent of human intestinal absorption (%HIA) is color-mapped as a function of  $\log P$  and/or  $pK_a$ . Quite obviously, the region with boundaries defined by acid  $pK_a$  ( $> 5$ ), and base  $pK_a$  ( $< 8$ ) shows as favorable for diffusion, while the  $\log P$  threshold for good absorption depends on the ionization state of the compound.

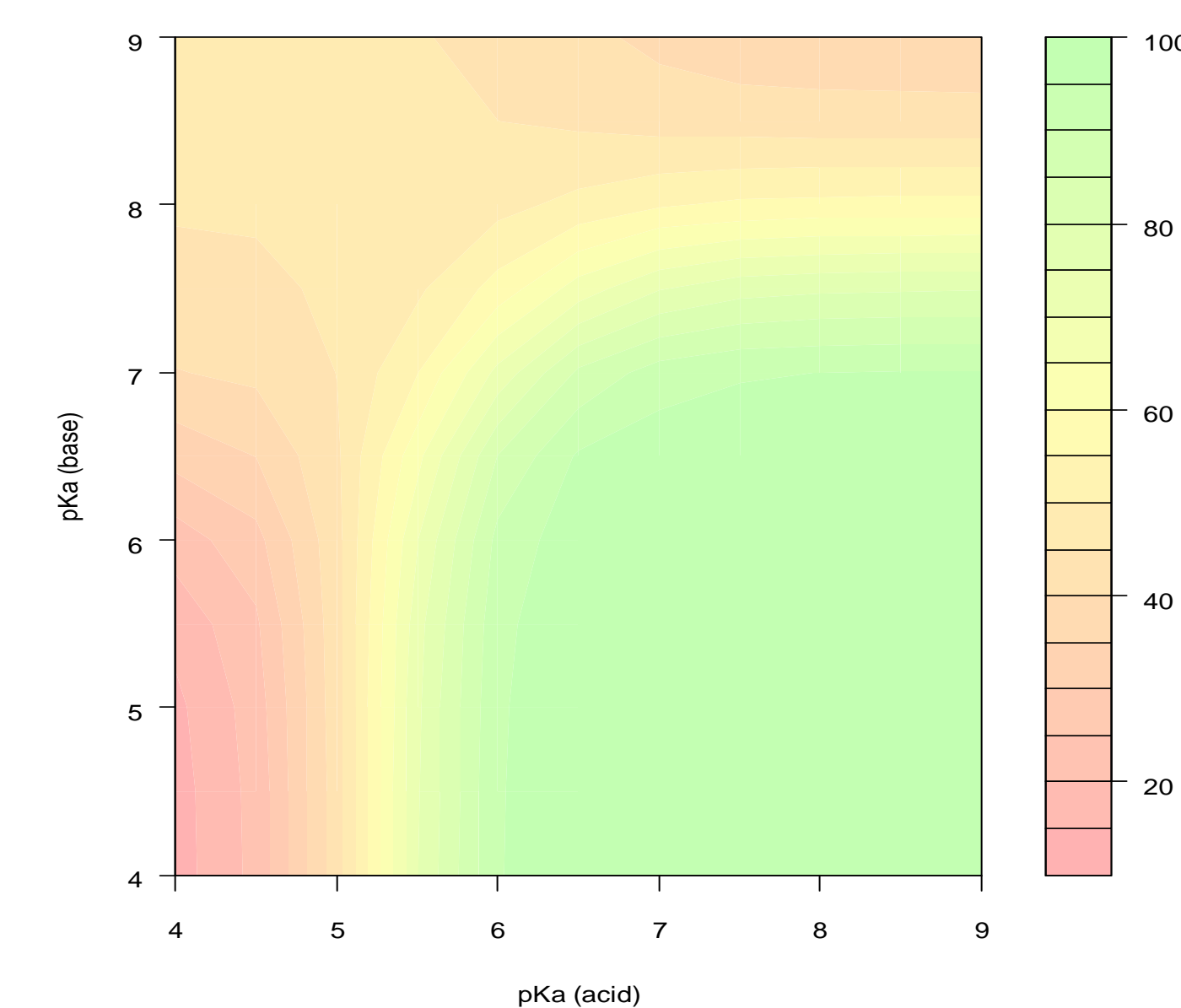


FIGURE 1. Contour plot of %HIA dependence on acid and base  $pK_a$  ( $\log P = 0.5$ ,  $N_{HD} = 3$ ).

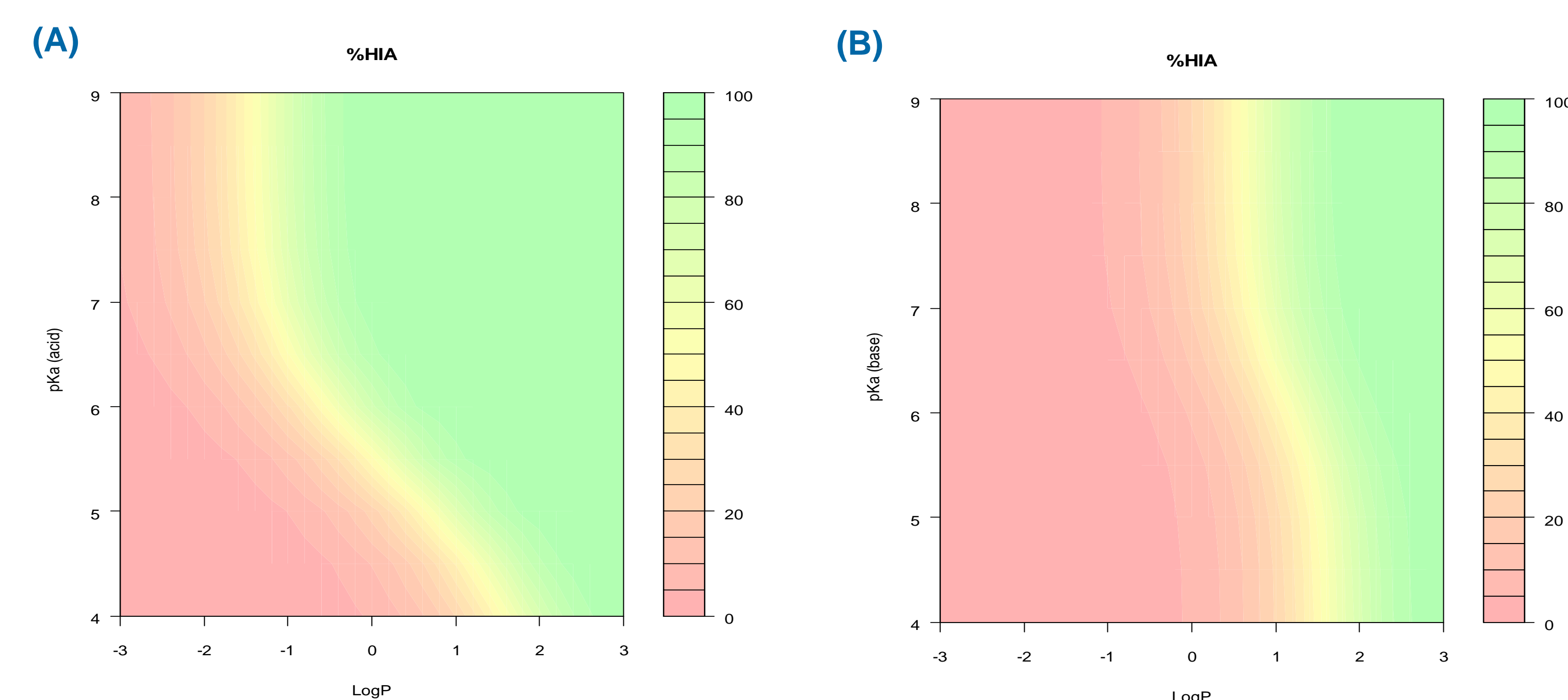


FIGURE 2. Contour plots of %HIA dependence on  $\log P$  and (A) acid  $pK_a$ ; (B) base  $pK_a$  ( $N_{HD} = 3$ ).

Simple physicochemical trends are evident for complex properties characterizing protein-ligand interactions. For example, a clear distinction can be made between substrates and non-substrates, as well as inhibitors and non-inhibitors of human P-glycoprotein (P-gp) on the basis of the compound's molecular weight, lipophilicity ( $\log P$ ), and H-bond accepting potential represented by Abraham's  $B$  parameter.<sup>5</sup>

### DEFINING OPTIMAL PHYSICOCHEMICAL SPACE

Given sets of physicochemical rules, it is possible to impose certain restrictions on the allowed parameter variation ranges and eliminate analogs falling into 'unfavorable' regions from further consideration. In our software each PhysChem parameter is evaluated for its contribution to various ADME properties. Again, consider the influence of  $\log P$  on %HIA. At each point of the  $\log P$  scale multiple %HIA predictions are performed with a fixed  $\log P$  value, and all other parameters allowed to vary within the ranges enclosed by sliders (Figure 3). The average of the predictions is color-mapped on the scale (red = 0%, green = 100%). If several ADME properties need to be optimized simultaneously, the software combines the output of relevant models with their 'weights' indicated by stars. The green intervals then define the regions in physicochemical space that are likely to produce analogs that would satisfy our project objectives.

Figure 3 (left) provides an example of lead optimization for CNS penetration (highest importance – 3 stars). Notably, a 'green zone' for  $\log P$  is identified in the 2-4 range, and the negative contribution of ionization is also evident. In Figure 3 (right) a peripheral drug is being designed and the target CNS score is reversed. Also, focus is brought to improving solubility (3 stars) and maintaining good absorption (1 star). As the opposite aims have now been set for BBB and intestinal permeability, the effects of most parameters are not as pronounced, and the major lead optimization direction would be decreasing lipophilicity.

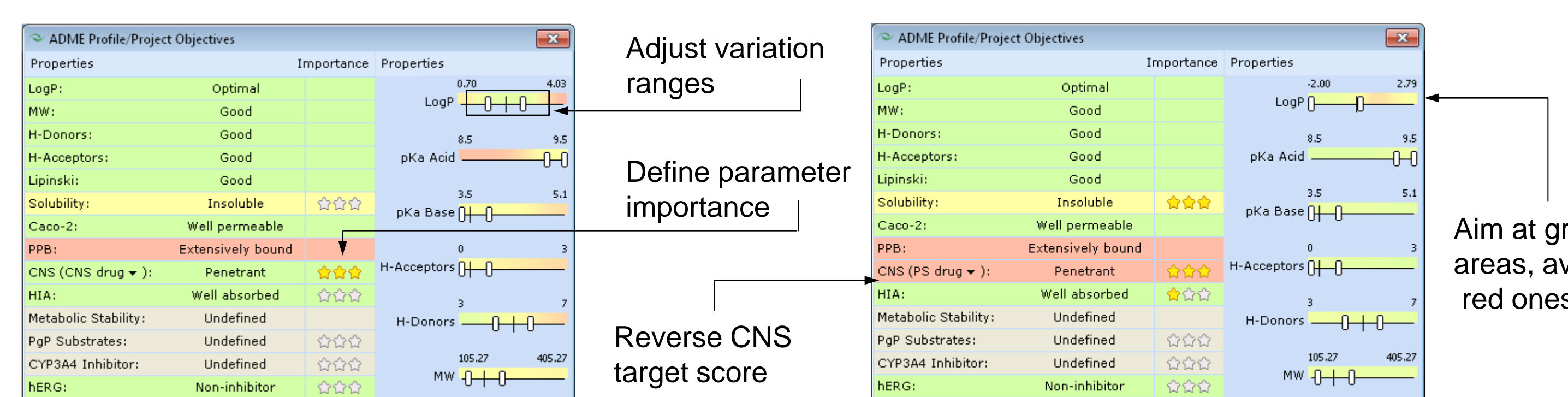


FIGURE 3. ADME Profile/Project Objectives—identifying favorable regions in physicochemical space.

### SUBSTITUENT SELECTION

Once the desired physicochemical profile of the target molecule is known, it has to be translated into a set of criteria for selecting suitable substituents to replace the fragment being optimized. In addition to simple molecular size, topology, and hydrogen bond count filters, our software calculates Hansch  $\pi$  to account for substituent contributions to the lipophilicity of the molecule.

Due to the significance of ionization for the majority of biological processes, a point of special interest is estimating  $pK_a$  of the ionizable groups in substituents connected to the given core scaffold, as well as the influence of these groups on the  $pK_a$  of the core molecule.  $\Delta pK_a$  of the core molecule caused by substituents can be calculated on the basis of its electronic parameters ( $\sigma_{inductive}$ ,  $\sigma_{resonance}$ ). Predicting  $pK_a$  of the substituent itself is accomplished by building custom Hammett-type equations for groups of model compounds containing the same ionizable group (Figure 4).

The Substituent panel depicted in Figure 4 is linked to ADME Profile/Project Objectives window, so that any changes in physicochemical restrictions for the target molecule are immediately reflected in the substituent selection criteria and vice versa.

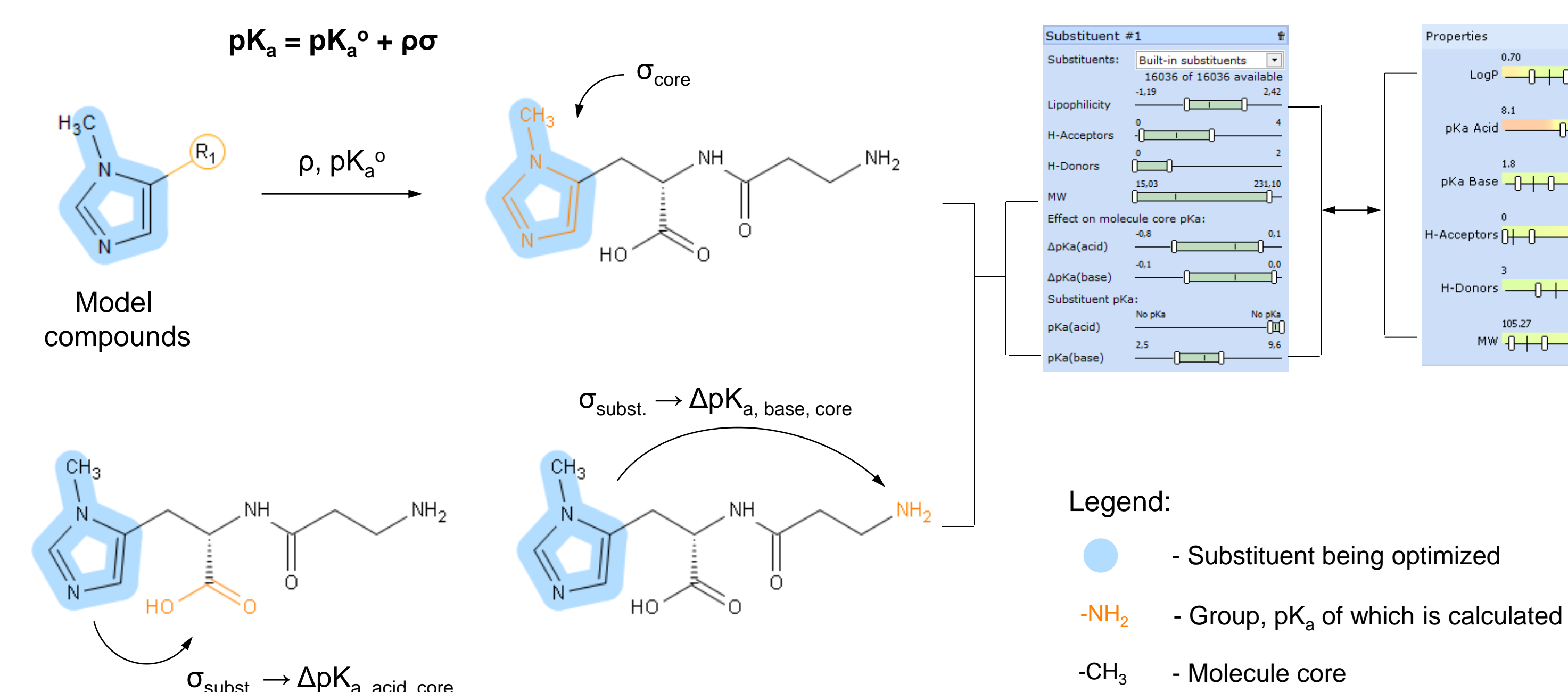


FIGURE 4. Estimating the influence of the substituent and core molecule on the analog's  $pK_a$ .

### COMPUTATIONAL PLATFORM FOR LEAD OPTIMIZATION

A variety of algorithms for calculating physicochemical, ADME and drug safety-related properties, as well as tools for generating, profiling and ranking analogs within selected regions of physicochemical space are available in the ACD/Percepta software platform as different workspaces (Figures 5–7):

#### 1. Structure Designer Workspace

The main interface for setting up project objectives, selecting substituents and generating analogs. Offers the following features:

- **Embedded structure editor** for quick in-place editing of the molecule to be optimized
- **ADME Profile/Project Objectives window**, described above. The predicted ADME profile is updated on the fly, as you draw
- **Substituent Panel** for fine-tuning the substituent properties and viewing the set of substituents that fulfill the established criteria

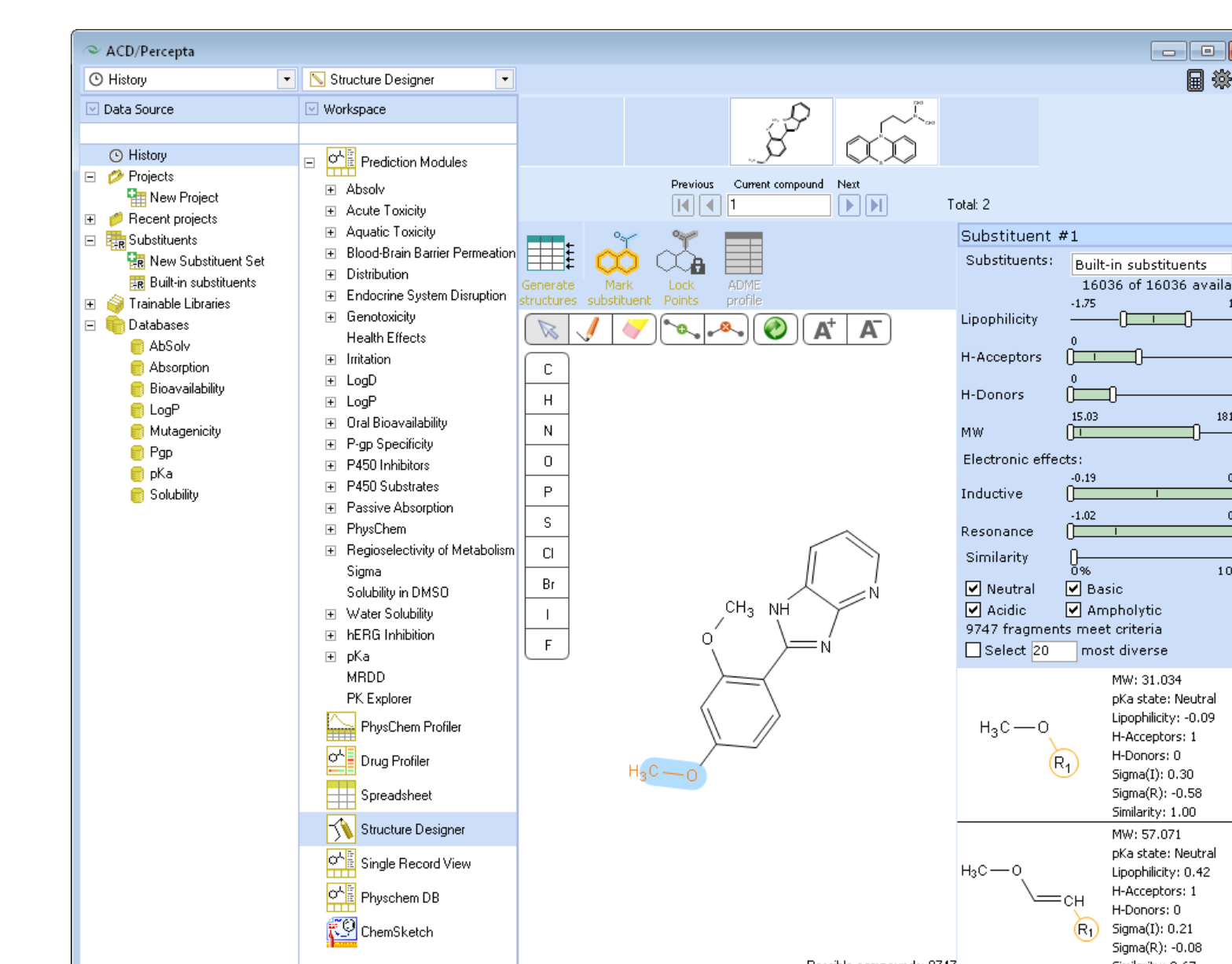


FIGURE 5. Structure Designer in ACD/Percepta.

#### 2. Spreadsheet Workspace

Provides advanced functionality for simultaneous analysis of multiple molecules:

- **Batch calculation** capabilities where property calculation and assignment of qualitative categories can be automatically performed for entire compound libraries including the ones generated by Structure Designer
- **Automatic ranking** according to a variety of predefined and user-supplied categories with the possibility to adjust the contribution of the individual properties on the overall rank score.
- **Plotting utilities** for graphical analysis of the relationships between different properties

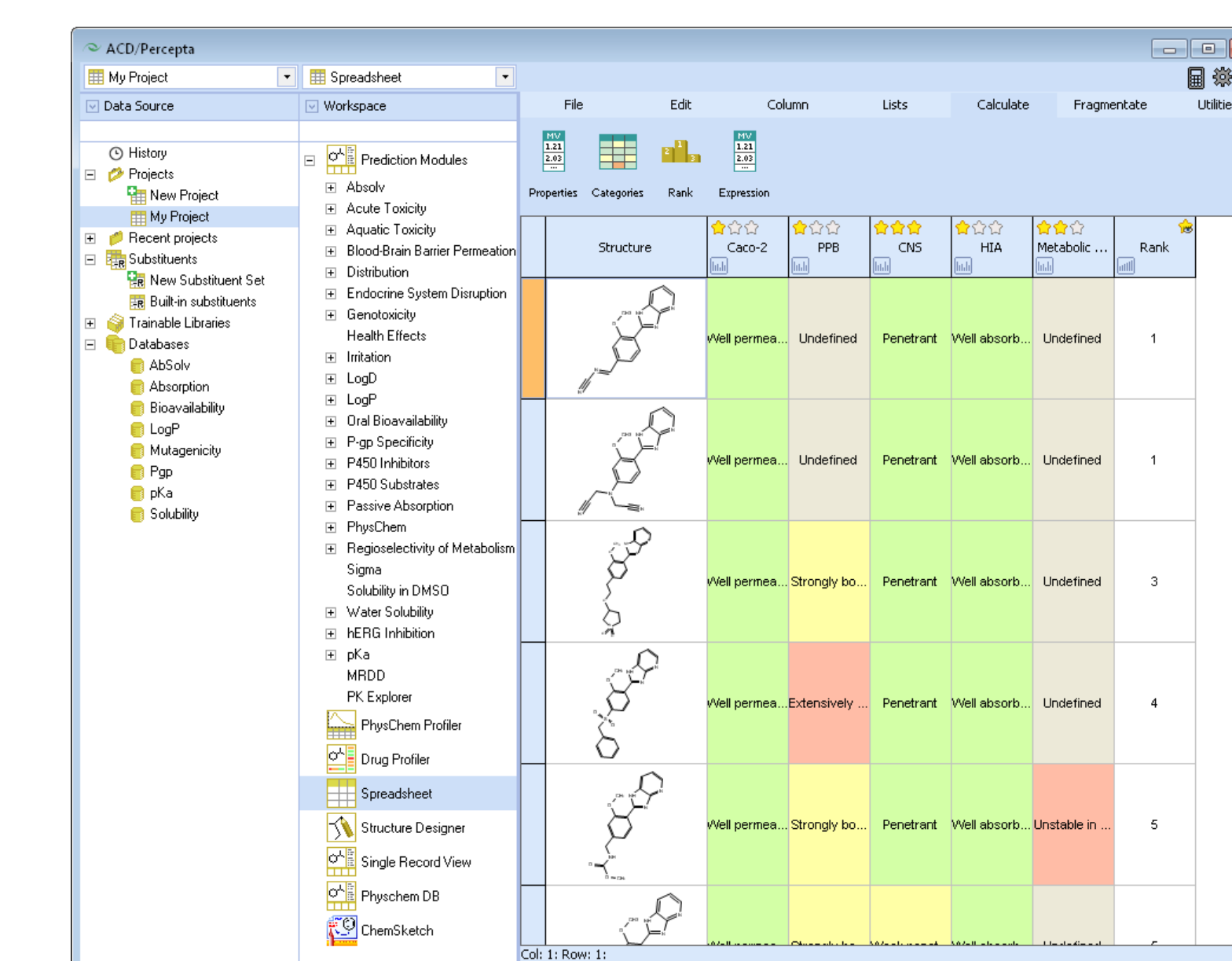


FIGURE 6. Spreadsheet workspace in ACD/Percepta.

#### 3. Prediction Workspace

- Dedicated **Prediction Modules** for particular properties provide additional comments and calculation details, as well as experimental values for similar compounds
- **Drug Profiler** module presents the compound's full ADME/Tox profile and enables customizing category assignment by employing user-defined property thresholds to differentiate between classes

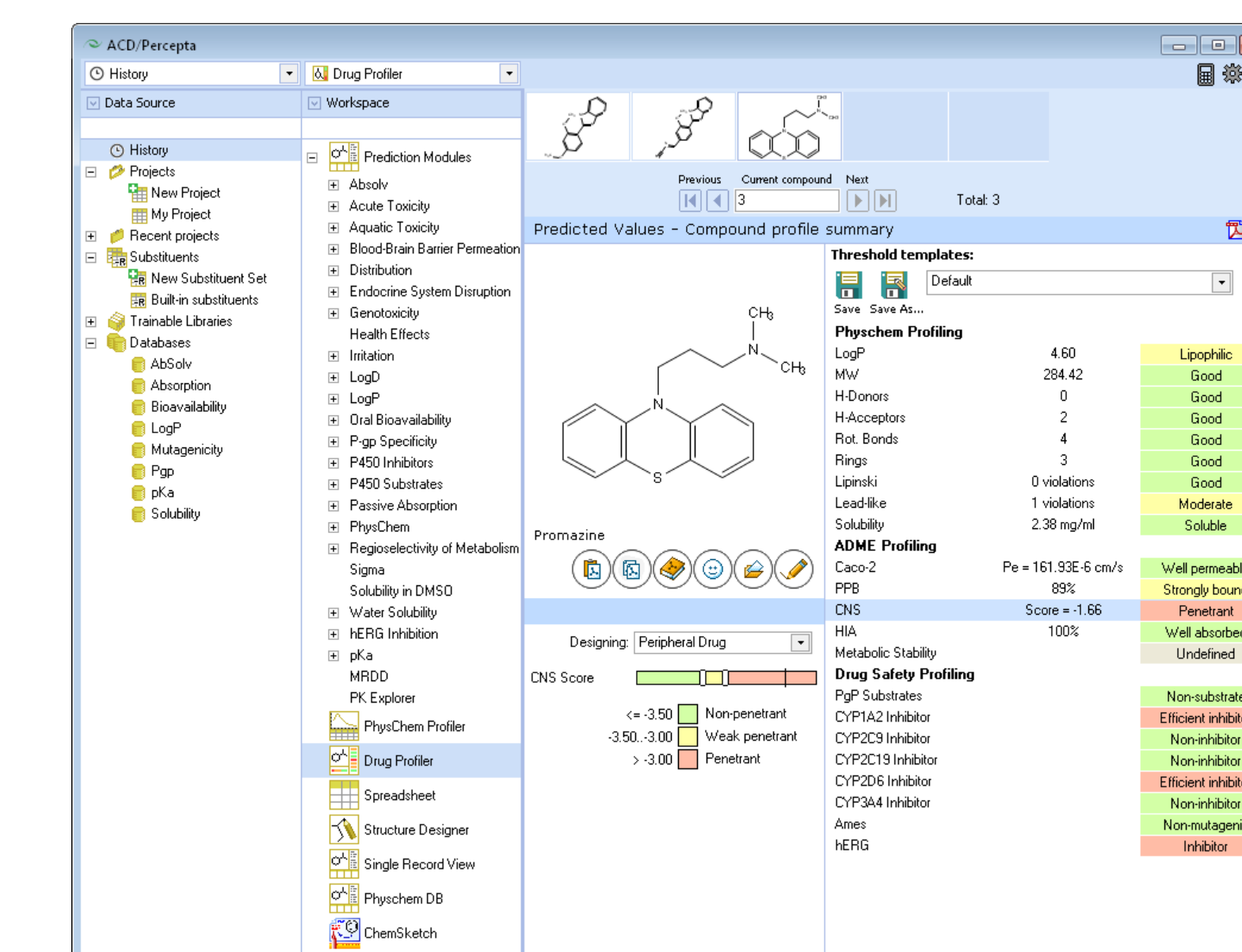


FIGURE 7. Drug Profiler module in ACD/Percepta.

### CONCLUSION

The presented property prediction and substituent selection methods coupled with automatic analog generation in accordance with the imposed physicochemical restrictions make our software platform a valuable tool to guide drug discovery projects towards the most promising candidates.

### REFERENCES

- [1] Reynolds D. P. et al. *J Pharm Sci.* **2009**, 98, 4039.
- [2] Lanevskij K. et al. *J Pharm Sci.* **2009**, 98, 122.
- [3] Lanevskij K. et al. *Chem Biodivers.* **2009**, 6, 2050.
- [4] Lanevskij K. et al. *J Pharm Sci.* **2011**, 100, 2147.
- [5] Didziapetris R. et al. *J Drug Target.* **2003**, 11, 391.

