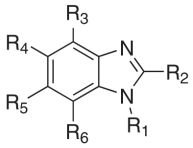
Supplementary information

# Structure-activity relationships of benzimidazole derivatives as antiparasitic agents: *Dual activity-difference (DAD) maps*

Jaime Pérez-Villanueva, Radleigh Santos, Alicia Hernández-Campos, Marc A. Giulianotti, Rafael Castillo, and Jose L. Medina-Franco

Table of Contents:

**Table S1.** Chemical structures of the benzimidazoles used in this study, and biological activity against *T. vaginalis* and *G. intestinalis*. The chemical representations **33**-**78**, include the two tautomeric forms of the 23 compounds. The biological activity of the two tautomeric forms is *approximated* to be the same.




| | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $pIC_{50}$ *T. vaginalis* | $pIC_{50}$ *G. intestinalis* | $\Delta pIC_{50}$ | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | H | $CF_3$ | H | H | H | H | 5.50 | 6.97 | -1.47 | 6 |
| **2** | $CH_3$ | $CF_3$ | H | $CF_3$ | H | H | 5.39 | 5.94 | -0.55 | 6 |
| **3** | $CH_3$ | $CF_3$ | H | H | $CF_3$ | H | 5.27 | 5.05 | 0.22 | 6 |
| **4** | $CH_3$ | $CF_3$ | H | Propylthio | H | H | 6.70 | 4.98 | 1.72 | 5 |
| **5** | $CH_3$ | $CF_3$ | H | H | Propylthio | H | 5.59 | 5.85 | -0.26 | 5 |
| **6** | $CH_3$ | $CF_3$ | H | Benzoyl | H | H | 4.53 | 5.96 | -1.43 | 5 |
| **7** | $CH_3$ | $CF_3$ | H | H | Benzoyl | H | 4.97 | 5.89 | -0.92 | 5 |
| **8** | H | $CF_3$ | H | Br | Br | H | 6.66 | 6.92 | -0.26 | 7 |
| **9** | H | $CF_3$ | Br | Br | Br | Br | 8.70 | 7.25 | 1.45 | 7 |
| **10** | H | $C_2F_5$ | H | Cl | Cl | H | 6.52 | 6.25 | 0.27 | 7 |
| **11** | H | $CF_3$ | H | $NO_2$ | $NO_2$ | H | 6.24 | 6.62 | -0.38 | 7 |
| **12** | H | $C_2F_5$ | Br | Br | Br | Br | 5.00 | 7.64 | -2.64 | 7 |
| **13** | $CH_3$ | $CONH_2$ | H | H | Cl | H | 6.96 | 7.12 | -0.16 | 9 |
| **14** | $CH_3$ | $CONHCH_3$ | H | H | Cl | H | 6.98 | 7.15 | -0.17 | 9 |
| **15** | $CH_3$ | $CON(CH_3)_2$ | H | H | Cl | H | 6.63 | 7.40 | -0.77 | 9 |
| **16** | $CH_3$ | $COOCH_2CH_3$ | H | H | Cl | H | 7.72 | 7.32 | 0.4 | 9 |
| **17** | $CH_3$ | $CONH_2$ | H | Cl | H | H | 6.73 | 6.63 | 0.1 | 9 |
| **18** | $CH_3$ | $CONHCH_3$ | H | Cl | H | H | 6.45 | 6.45 | 0 | 9 |
| **19** | $CH_3$ | $CON(CH_3)_2$ | H | Cl | H | H | 6.68 | 6.61 | 0.07 | 9 |
| **20** | $CH_3$ | $COOCH_2CH_3$ | H | Cl | H | H | 7.57 | 7.40 | 0.17 | 9 |
| **21** | $CH_3$ | $CONH_2$ | H | Cl | Cl | H | 6.87 | 6.34 | 0.53 | 9 |
| **22** | $CH_3$ | $CONHCH_3$ | H | Cl | Cl | H | 6.65 | 6.82 | -0.17 | 9 |
| **23** | $CH_3$ | $CON(CH_3)_2$ | H | Cl | Cl | H | 7.12 | 7.13 | -0.01 | 9 |
| **24** | $CH_3$ | $COOCH_2CH_3$ | H | Cl | Cl | H | 7.53 | 7.56 | -0.03 | 9 |
| **25** | $CH_3$ | $CONH_2$ | H | H | H | H | 6.78 | 7.03 | -0.25 | 9 |
| **26** | $CH_3$ | $CONHCH_3$ | H | H | H | H | 6.98 | 7.22 | -0.24 | 9 |
| **27** | $CH_3$ | $CON(CH_3)_2$ | H | H | H | H | 6.37 | 6.29 | 0.08 | 9 |
| **28** | $CH_3$ | $COOCH_2CH_3$ | H | H | H | H | 7.07 | 7.16 | -0.09 | 9 |
| **29** | $CH_3$ | $COCH_3$ | H | H | H | H | 6.68 | 7.06 | -0.38 | 9 |
| **30** | $CH_3$ | $COCH_3$ | H | Cl | H | H | 6.88 | 7.30 | -0.42 | 9 |
| **31** | $CH_3$ | $COCH_3$ | H | H | Cl | H | 6.64 | 7.17 | -0.53 | 9 |
| **32** | $CH_3$ | $COCH_3$ | H | Cl | Cl | H | 7.20 | 7.46 | -0.26 | 9 |
| **33** | H | $CF_3$ | H | Cl | H | H | 6.35 | 5.89 | 0.46 | 6 |
| **34** | H | $CF_3$ | H | H | Cl | H | 6.35 | 5.89 | 0.46 | 6 |
| **35** | H | $CF_3$ | H | F | H | H | 5.50 | 6.31 | -0.81 | 6 |
| **36** | H | $CF_3$ | H | H | F | H | 5.50 | 6.31 | -0.81 | 6 |

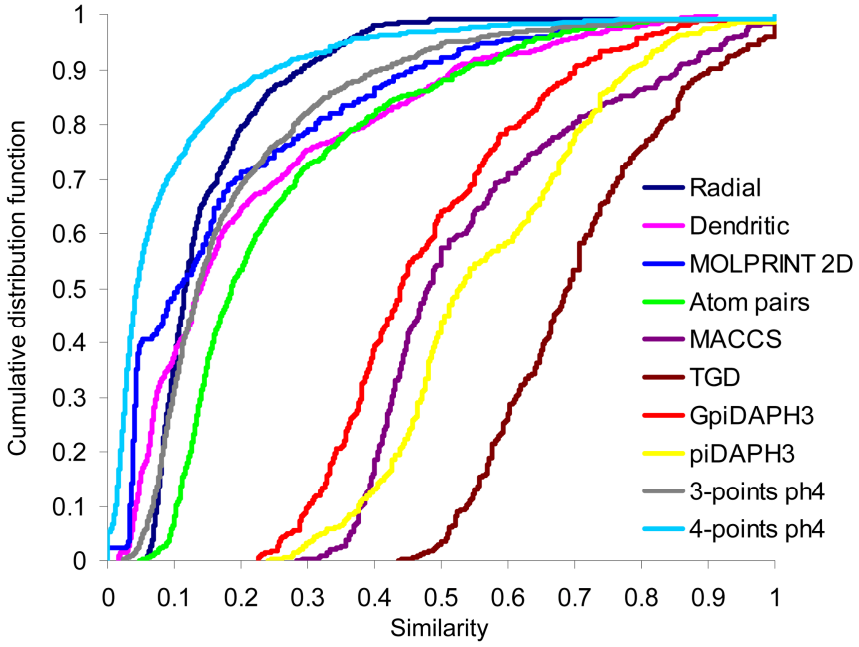
**Table S1.** (continued)

| | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $pIC_{50}$ *T. vaginalis* | $pIC_{50}$ *G. intestinalis* | $\Delta pIC_{50}$ | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|
| **37** | H | $CF_3$ | H | $CF_3$ | H | H | 6.63 | 6.17 | 0.46 | 6 |
| **38** | H | $CF_3$ | H | H | $CF_3$ | H | 6.63 | 6.17 | 0.46 | 6 |
| **39** | H | $CF_3$ | H | CN | H | H | 5.64 | 5.75 | -0.11 | 6 |
| **40** | H | $CF_3$ | H | H | CN | H | 5.64 | 5.75 | -0.11 | 6 |
| **41** | H | $CF_3$ | H | Propiltio | H | H | 6.46 | 5.82 | 0.64 | 5 |
| **42** | H | $CF_3$ | H | H | Propiltio | H | 6.46 | 5.82 | 0.64 | 5 |
| **43** | H | $CF_3$ | H | Benzoil | H | H | 4.55 | 5.64 | -1.09 | 5 |
| **44** | H | $CF_3$ | H | H | Benzoil | H | 4.55 | 5.64 | -1.09 | 5 |
| **45** | H | $CF_3$ | H | Br | H | H | 5.80 | 6.19 | -0.39 | 7 |
| **46** | H | $CF_3$ | H | H | Br | H | 5.80 | 6.19 | -0.39 | 7 |
| **47** | H | $CF_3$ | Br | H | Br | H | 6.72 | 6.60 | 0.12 | 7 |
| **48** | H | $CF_3$ | H | Br | H | Br | 6.72 | 6.60 | 0.12 | 7 |
| **49** | H | $CF_3$ | Br | Br | Br | H | 6.57 | 6.85 | -0.28 | 7 |
| **50** | H | $CF_3$ | H | Br | Br | Br | 6.57 | 6.85 | -0.28 | 7 |
| **51** | H | $SCH_2CH_2OH$ | Cl | H | Cl | H | 6.16 | 7.42 | -1.26 | 8 |
| **52** | H | $SCH_2CH_2OH$ | H | Cl | H | Cl | 6.16 | 7.42 | -1.26 | 8 |
| **53** | H | $SCH_2CH_2OH$ | Br | H | Br | H | 6.83 | 6.98 | -0.15 | 8 |
| **54** | H | $SCH_2CH_2OH$ | H | Br | H | Br | 6.83 | 6.98 | -0.15 | 8 |
| **55** | H | $SCH_2CH_2N(CH_3)_2$ | Cl | H | Cl | H | 7.14 | 7.68 | -0.54 | 8 |
| **56** | H | $SCH_2CH_2N(CH_3)_2$ | H | Cl | H | Cl | 7.14 | 7.68 | -0.54 | 8 |
| **57** | H | $SCH_2CH_2N(CH_3)_2$ | Br | H | Br | H | 6.51 | 7.54 | -1.03 | 8 |
| **58** | H | $SCH_2CH_2N(CH_3)_2$ | H | Br | H | Br | 6.51 | 7.54 | -1.03 | 8 |
| **59** | H | $SCH_2CH_2N(C_2H_5)_2$ | Cl | H | Cl | H | 7.39 | 7.72 | -0.33 | 8 |
| **60** | H | $SCH_2CH_2N(C_2H_5)_2$ | H | Cl | H | Cl | 7.39 | 7.72 | -0.33 | 8 |
| **61** | H | $SCH_2CH_2N(C_2H_5)_2$ | Br | H | Br | H | 7.53 | 7.46 | 0.07 | 8 |
| **62** | H | $SCH_2CH_2N(C_2H_5)_2$ | H | Br | H | Br | 7.53 | 7.46 | 0.07 | 8 |
| **63** | H | $SCH_2CH_2CH_2N(CH_3)_2$ | Cl | H | Cl | H | 7.07 | 7.14 | -0.07 | 8 |
| **64** | H | $SCH_2CH_2CH_2N(CH_3)_2$ | H | Cl | H | Cl | 7.07 | 7.14 | -0.07 | 8 |
| **65** | H | $SCH_2CH_2CH_2N(CH_3)_2$ | Br | H | Br | H | 7.99 | 7.69 | 0.3 | 8 |
| **66** | H | $SCH_2CH_2CH_2N(CH_3)_2$ | H | Br | H | Br | 7.99 | 7.69 | 0.3 | 8 |
| **67** | H | $SCH_2CH_2$-(*N*-piperidyl) | Cl | H | Cl | H | 7.14 | 7.26 | -0.12 | 8 |
| **68** | H | $SCH_2CH_2$-(*N*-piperidyl) | H | Cl | H | Cl | 7.14 | 7.26 | -0.12 | 8 |
| **69** | H | $SCH_2CH_2$-(*N*-piperidyl) | Br | H | Br | H | 7.69 | 6.90 | 0.79 | 8 |
| **70** | H | $SCH_2CH_2$-(*N*-piperidyl) | H | Br | H | Br | 7.69 | 6.90 | 0.79 | 8 |
| **71** | H | $SCH_2CH_2$-(*N*-morpholinyl) | Cl | H | Cl | H | 7.29 | 7.41 | -0.12 | 8 |
| **72** | H | $SCH_2CH_2$-(*N*-morpholinyl) | H | Cl | H | Cl | 7.29 | 7.41 | -0.12 | 8 |
| **73** | H | $SCH_2CH_2$-(*N*-morpholinyl) | Br | H | Br | H | 7.23 | 7.23 | 0 | 8 |
| **74** | H | $SCH_2CH_2$-(*N*-morpholinyl) | H | Br | H | Br | 7.23 | 7.23 | 0 | 8 |
| **75** | H | $SCH_2CH_2$-(*p*-nitrophenyl) | Cl | H | Cl | H | 7.96 | 7.09 | 0.87 | 8 |
| **76** | H | $SCH_2CH_2$-(*p*-nitrophenyl) | H | Cl | H | Cl | 7.96 | 7.09 | 0.87 | 8 |
| **77** | H | $SCH_2CH_2$-(*p*-nitrophenyl) | Br | H | Br | H | 8.48 | 7.43 | 1.05 | 8 |
| **78** | H | $SCH_2CH_2$-(*p*-nitrophenyl) | H | Br | H | Br | 8.48 | 7.43 | 1.05 | 8 |

**Table S2.** Correlation matrix for the pairwise molecular similarities.

| | Radial | Dendritic | MOLPRINT 2D | Atom pairs | MACCS | TGD | GpiDAPH3 | piDAPH3 | 3-point ph4 | 4-point ph4 |
|---|---|---|---|---|---|---|---|---|---|---|
| Radial | 1.00 | | | | | | | | | |
| Dendritic | 0.72 | 1.00 | | | | | | | | |
| MOLPRINT 2D | 0.87 | 0.80 | 1.00 | | | | | | | |
| Atom pairs | 0.87 | 0.82 | 0.93 | 1.00 | | | | | | |
| MACCS | 0.72 | 0.74 | 0.78 | 0.88 | 1.00 | | | | | |
| TGD | 0.56 | 0.67 | 0.58 | 0.70 | 0.69 | 1.00 | | | | |
| GpiDAPH3 | 0.59 | 0.74 | 0.66 | 0.72 | 0.65 | 0.68 | 1.00 | | | |
| piDAPH3 | 0.55 | 0.69 | 0.61 | 0.65 | 0.56 | 0.64 | 0.91 | 1.00 | | |
| 3-point ph4[a] | 0.63 | 0.77 | 0.70 | 0.75 | 0.71 | 0.68 | 0.78 | 0.75 | 1.00 | |
| 4-point ph4[a] | 0.60 | 0.71 | 0.65 | 0.70 | 0.65 | 0.60 | 0.72 | 0.69 | 0.96 | 1.00 |

[a] ph4: pharmacophore




| Representation | Max | Q3 | Median | Q1 | Min | Mean | STD |
|---|---|---|---|---|---|---|---|
| Radial | 1.00 | 0.18 | 0.12 | 0.09 | 0.05 | 0.16 | 0.12 |
| Dendritic | 0.91 | 0.30 | 0.14 | 0.07 | 0.02 | 0.22 | 0.20 |
| MOLPRINT 2D | 1.00 | 0.25 | 0.11 | 0.04 | 0.00 | 0.18 | 0.19 |
| Atom pairs | 1.00 | 0.33 | 0.19 | 0.13 | 0.05 | 0.26 | 0.18 |
| MACCS | 1.00 | 0.64 | 0.48 | 0.42 | 0.28 | 0.55 | 0.18 |
| TGD | 1.00 | 0.80 | 0.69 | 0.59 | 0.44 | 0.70 | 0.14 |
| GpiDAPH3 | 1.00 | 0.58 | 0.44 | 0.36 | 0.23 | 0.48 | 0.16 |
| piDAPH3 | 1.00 | 0.69 | 0.53 | 0.46 | 0.24 | 0.57 | 0.16 |
| 3-points ph4 | 1.00 | 0.24 | 0.14 | 0.09 | 0.02 | 0.20 | 0.17 |
| 4-points ph4 | 1.00 | 0.12 | 0.04 | 0.02 | 0.00 | 0.10 | 0.15 |

**Figure S1.** Cumulative distribution functions of 3,003 pairwise molecular similarities using different 2D and 3D fingerprint representations. The table summarizes the information of the distributions. Q3 and Q1 indicate the third and first quartile, respectively.

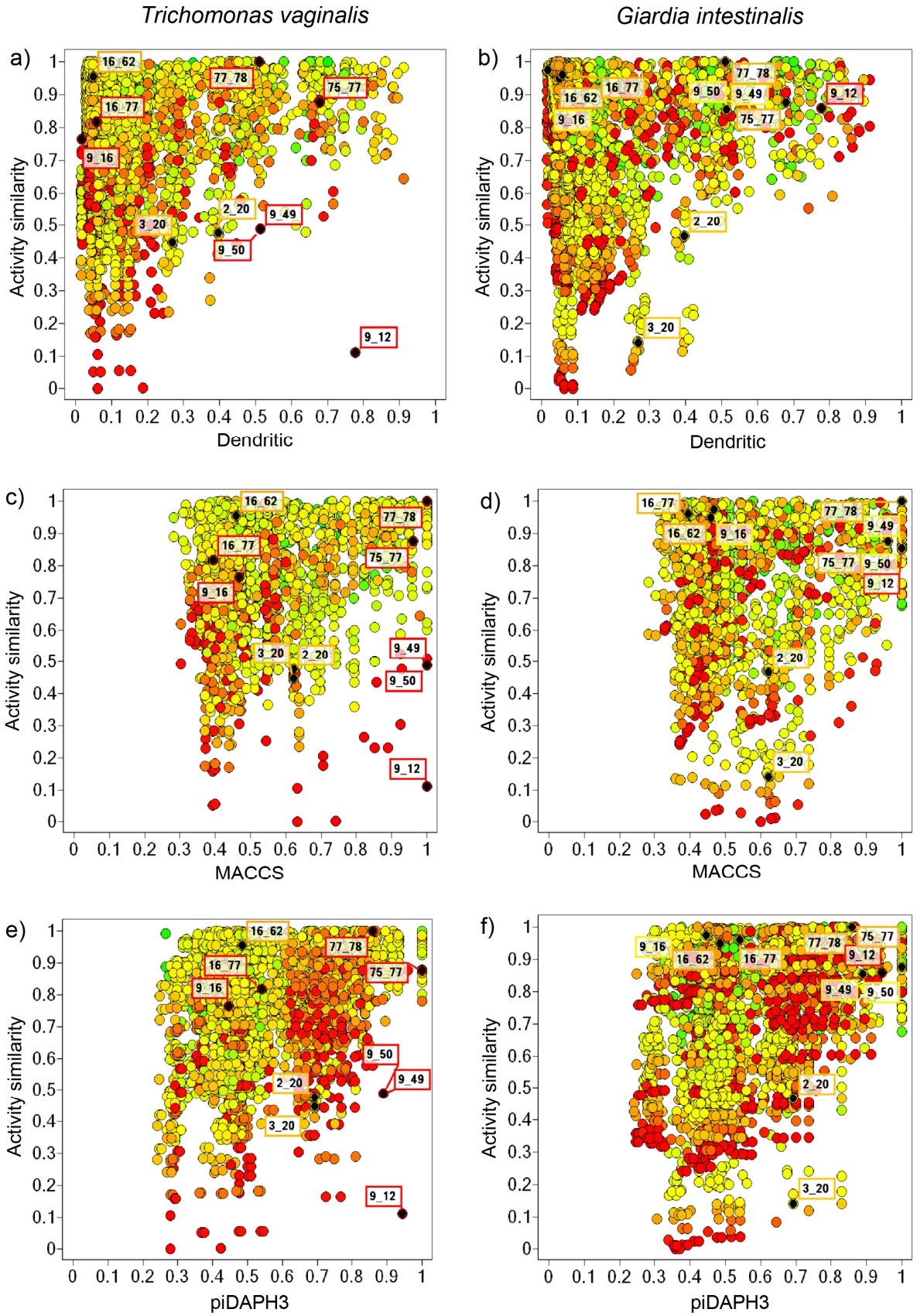



**Figure S2.** SAS maps for *T. vaginalis* and *G. intestinalis* with different structural representations: (a) and (b) Dendritic; (c) and (d) MACCS keys; (e) and (f) piDAPH3. Data points are color-coded by the activity of the most active compound in the pair using a continuous scale from green (less active) to red (most active). The figure using a continuous scale from white to black is shown in Figure S5.

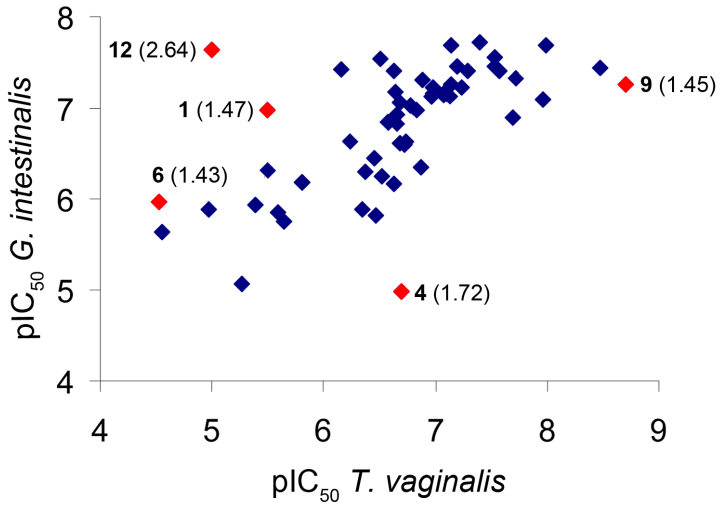



**Figure S3.** Correlation between the $pIC_{50}$ values for *T. vaginalis* and *G. intestinalis*. Five compounds with the largest potency difference are marked in red. The absolute value of potency difference ($|\Delta pIC_{50}|$) is indicated in parenthesis. On average, the $pIC_{50}$ values for *T. vaginalis* and *G. intestinalis* have a percentage error of 10.1% (based on the error information available for compounds **1**-**3**, **8**-**40** and **45**-**78**).

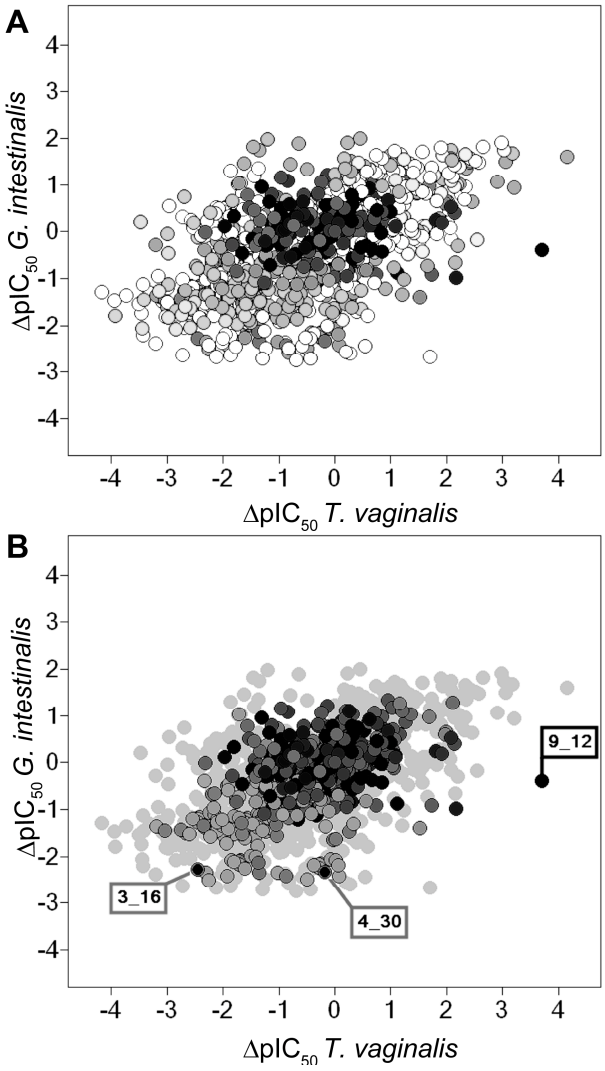



**Figure S4.** (a) DAD map with data pairs color-coded by dendritic fingerprint similarity using a continuous scale from white (less similar) to black (more similar). (b) Map showing in dark gray-to-black 925 data pairs with high similarity. Selected pairs are labeled with the compound numbers. The chemical structures of compounds in these pairs are shown in Figure 4.

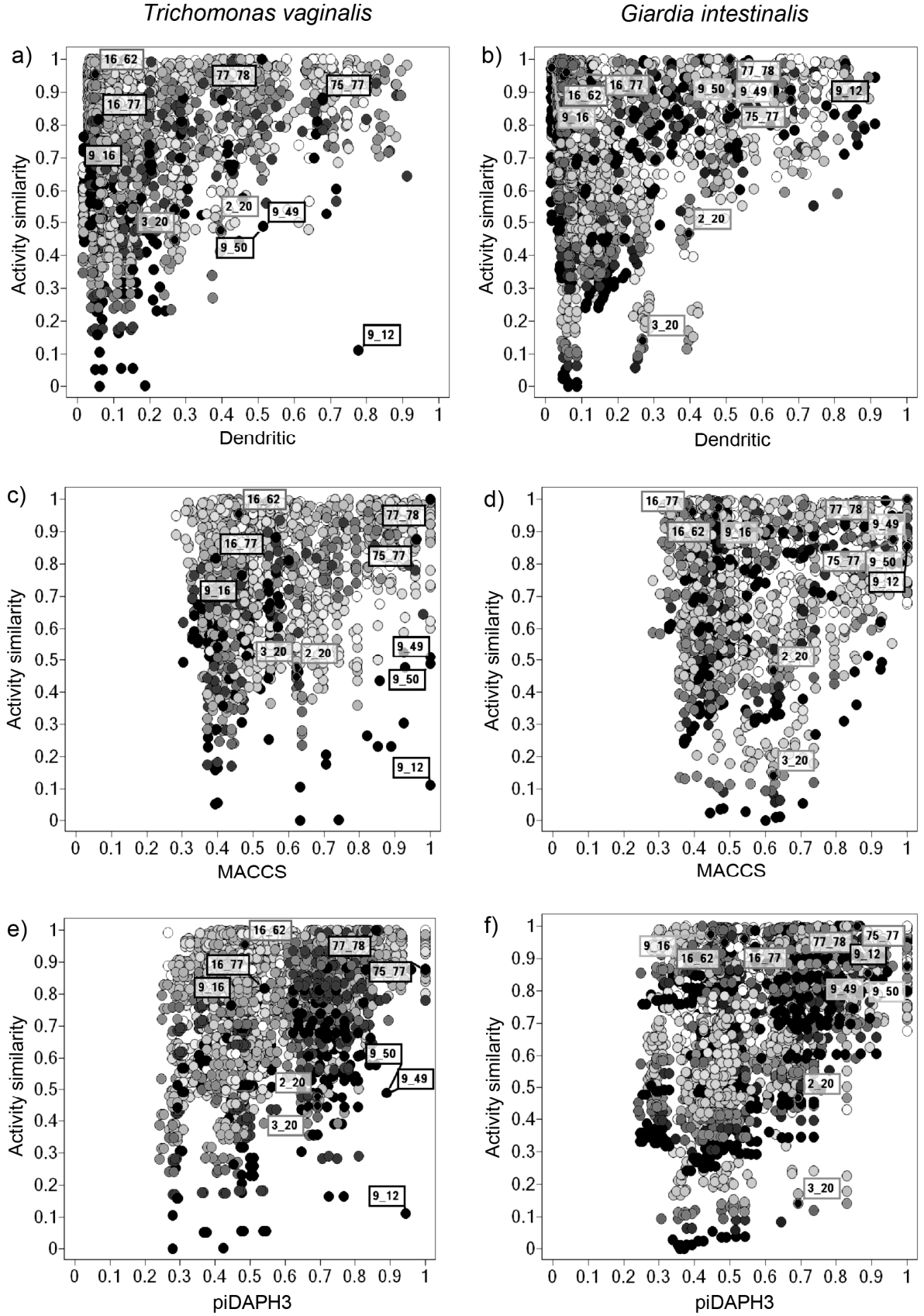



**Figure S5.** SAS maps for *T. vaginalis* and *G. intestinalis* with different structural representations: (a) and (b) Dendritic; (c) and (d) MACCS keys; (e) and (f) piDAPH3. Data points are color-coded by the activity of the most active compound in the pair using a continuous scale from white (less active) to black (most active).