Systematic Study of Effects of Structural Modifications on the Aqueous Solubility of Drug-like Molecules

José A. Cisneros, Michael J. Robertson, Brandon Q. Mercado, and William L. Jorgensen^{*@}

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States

Supporting Information

ABSTRACT: Aqueous solubilities and activities have been measured for 17 members of the quinolinyltriazole series of inhibitors of human macrophage migration inhibitory factor (MIF). Systematic variation of a solvent-exposed substituent provided increases in solubility from 2 $\mu\text{g/mL}$ for the parent compound **3a** up to 867 $\mu\text{g/mL}$. The low solubility of **3a** results from its near-planar structure and an intermolecular hydrogen bond, as revealed in a small-molecule X-ray structure. Removal of the hydrogen bond yields a 3-fold increase in solubility, but a 7-fold drop in activity. **5b** emerges as the most potent MIF inhibitor with a K_i of 14 nM and good solubility, 47 $\mu\text{g/mL}$, while **4e** has both high potency and solubility.

KEYWORDS: Aqueous solubility, MIF inhibitors, crystallography

Aqueous solubility is well-known as a critical property in the development of drugs.^{1–5} Poor solubility is associated with difficulties in the reliable performance of assays, in obtaining oral formulations for *in vivo* administration, and in bioavailability. On the other hand, high solubility may lead to poor cell permeability and rapid excretion. The optimal range for the aqueous solubility *S* of drugs intended for oral delivery is ca. 10 μM to 10 mM or, equivalently, 4–4000 $\mu\text{g/mL}$ for a compound with a molecular weight of 400.² Oral drugs with solubilities below 1 μM are very rare.

Poor solubility is a common issue in drug discovery efforts, since hydrophobic compounds bind well to target proteins for the same reasons that proteins fold to shield their hydrophobic groups from the aqueous environment. The problems may arise early from poorly soluble compounds that are obtained as hits in high-throughput screening. They often become aggravated during lead optimization, since the most reliable way to increase potency is to add hydrophobic substituents that fill hydrophobic regions in the target's binding site. Consistently, target proteins with highly hydrophobic binding sites are particularly prone to inhibitor designs with poor solubility. A classic example is non-nucleoside inhibitors of HIV-1 reverse transcriptase, NNRTIs.⁶ The binding site in this case features a cluster of residues with aliphatic and aromatic side chains, and poor solubility has characterized many classes of NNRTIs. This is especially true for diaminopyrimidines, including the FDA-approved drugs etravirine and rilpivirine (**1**). In comparison to **1**, we were able to increase the solubility ca. 700-fold to 14 $\mu\text{g/mL}$ for the triazine analogue with a morpholinylpropoxy substituent (**2**) while retaining excellent potency in infected T-cell assays.⁷ This illustrates the basic principle for improving solubility without loss of potency: add conformationally flexible substituents with polar groups to a site in the inhibitor that is solvent-exposed in the complex with the protein. The benefit of flexibility includes the entropic gain from populating more conformers in solution than

in the crystal. A corollary is to reduce planarity for inhibitors with multiple aromatic rings, which also leads to less tight packing in the crystalline state.⁸ However, the addition of polar groups comes with uncertainties, since they may also form stabilizing hydrogen bonds in the crystal.

More recently, solubility has become an issue in our development of inhibitors of the tautomerase activity of the human macrophage migration inhibitory factor (MIF).⁹ Specifically, surprisingly low solubility of 2.2 $\mu\text{g/mL}$ was found for the parent quinolinyltriazole **3a** ($R = \text{H}$), which arose from *de novo* design. However, modeling indicated that substituents at the 6- and 7-positions of the quinoline ring should be solvent exposed. This expectation was confirmed by obtaining X-ray

Received: November 8, 2016

Accepted: December 1, 2016

Published: December 1, 2016

crystal structures for **3a** and the R = methoxyethoxy (MOEO) analogue **3d** (Figure 1). The binding site is again seen to include

Figure 1. Rendering from the 1.8-Å crystal structure of **3d** bound to MIF (PDB ID: 4WRB).⁹ Carbon atoms of **3d** are colored yellow. Some residues in front of the ligand have been removed for clarity. Coordination of Lys32A is highlighted with dashed lines; the methoxyethoxy group on C6 of the quinoline is solvent-exposed.

multiple residues with aromatic and aliphatic side chains. The poor solubility of **3a** in spite of its four nitrogen atoms and hydroxyl group can be attributed to its expected near planarity,¹⁰ which has now been confirmed by a small molecule crystal structure.¹¹ As illustrated in Figure 2, the nearly planar monomers are well-stacked in the crystal structure and there are also intermolecular hydrogen bonds between the quinoline nitrogen atoms and phenolic hydroxyl groups (2.73 Å) in adjacent molecules (Figure 3). As expected, **3a** in isolation or in the crystal adopts a conformation with the quinoline nitrogen atom *anti* to N3 of the triazole ring to minimize lone-pair repulsion, while the conformation is *syn* in the complex with MIF to provide optimal coordination of the ammonium group of Lys32A (Figure 1).

Under the circumstances, it was decided to explore the solubility range that could be obtained by systematic variation of the substituent R in the 6-position of **3** and **4**. Two analogues of **5**⁹ were also considered along with the deshydroxy compound **6**. Disruption of planarity by, e.g., addition of a substituent at C3 of

Figure 2. Illustration of the packing in the 0.84-Å crystal structure of **3a**. The space group is $Pca2_1$. The thermal ellipsoids are depicted at the 50% probability level. The CCDC ID is 1514977.

Figure 3. Close-up of the intermolecular hydrogen bonding in the crystal structure of **3a**.

the quinoline is not viable, as it is accompanied by large decreases in potency in view of the slot-like binding site (Figure 1) and coordination of Lys32.⁹ In all, aqueous solubilities and inhibition constants K_i were determined for 17 analogues with 11 different substituents, as reported in Table 1. The aqueous solubilities were measured with a standard shake-flask procedure.^{6,9,12} Saturated solutions are obtained by stirring for 2 days in Britton–Robinson buffer (pH 6.5), followed by filtration (Acrodisc syringe, 0.2 μ m pore) and UV–vis analysis (Agilent 8453). Piroxicam has been used as a control more than 10 times, yielding $S = 6.5 \pm 1.7 \mu\text{g/mL}$, which is consistent with a reference value of $6.36 \pm 0.04 \mu\text{g/mL}$.¹² The inhibition constants were also determined as before using 4-hydroxyphenylpyruvic acid (HPP) as the substrate.^{9,13,14} Inhibitory activity is monitored by measuring formation of the borate complex of the enol product at 305 nm using a Tecan Infinite F500 plate reader. Nine of the 17 inhibitors have been reported previously;^{9,13} the synthetic and spectroscopic details for the new compounds (**3b**, **3c**, **3e**, **3f**, **3i**, **4c**, **4e**, **6**) are provided in the Supporting Information.

As reflected in Table 1, to improve solubility, we favor substituents that contain alkyleneoxy linkages with hydroxy, amino, and carboxylic acid termini. Amide and ester components are viewed as less desirable for drug prospects, owing to their potential decomposition by proteolytic enzymes. All compounds showed solubility gains over the parent **3a**. Among the smallest substituents, the benefits for the hydroxyethoxy and aminoethoxy analogues **3b** and **3c** were modest, while the carboxymethoxy analogue **3i** provided a striking solubility boost to 365 $\mu\text{g/mL}$.

Though it worked well in the present cases, in general, the impact of addition of substituents with a carboxylic acid is uncertain,

Table 1. Computed Octanol/Water log P's, Experimental Aqueous Solubility at pH 6.5 (*S* in $\mu\text{g/mL}$), and K_i (μM)

Cp	R	ClogP	QP ^a	S	K_i
3a	H	3.80	3.32	2.2	0.23
3b	HOCH ₂ CH ₂ O	3.22	2.88	2.6	0.53
3c ^b	H ₂ NCH ₂ CH ₂ O	3.29	2.12	3.7	0.26
3d	H ₃ COCH ₂ CH ₂ O	3.98	3.49	3.6	0.20
3e	H ₃ CO(CH ₂ CH ₂ O) ₂	3.85	3.81	2.4	0.147
3f	2-THP-CH ₂ O	4.89	4.07	3.4	0.27
3g ^b	H ₂ N(CH ₂ CH ₂ O) ₂	3.36	2.43	13.9	0.36
3h ^c	4-Mr(CH ₂ CH ₂ O) ₂	4.24	2.82	48.5	0.161
3i	HOOCCH ₂ O	3.04	2.64	365	0.20
4a ^b	H ₂ N(CH ₂ CH ₂ O) ₂	3.35	2.77	9.1	0.144
4b	4-Mr(CH ₂ CH ₂ O) ₂	4.24	3.16	27.2	0.074
4c	HOOCCH ₂ O	3.38	2.98	37.0	0.048
4d	HOOC(CH ₂) ₃ O	4.06	3.90	19.2	0.039
4e	HOOCCH ₂ OCH ₂ CH ₂ O	3.63	3.30	867	0.037
5a	H ₃ COCH ₂ CH ₂ O	5.70	5.54	6.1	0.024
5b	HOOC	5.64	4.70	47.2	0.014
6 ^d	HOOCCH ₂ O	2.48	3.32	1046	1.37
7 ^e					27.3

^aQPlogP. ^bTFA salt. ^cMr = morpholinyl. ^ddes-Hydroxy analogue of 3i. ^e(R)-ISO-1.

since carboxylic acids often form hydrogen-bonded dimers in their crystals.^{4,15} The purely ether-containing substituents in **3d**, **3e**, and **3f** also just gave modest improvements for the solubility, and it is notable that addition of a second ethyleneoxy unit in going from **3d** to **3e** was not helpful. However, significant gains were found with amino-containing substituents in **3g** and **3h**.¹⁶ It is well-known that introduction of morpholine and related heterocycles is normally successful in improving aqueous solubility.^{5,7} The groups are expected to be protonated at physiological pH, which benefits hydration, and their saturated, nonplanar character eschews tight crystal packing.

Turning to **4a–4e**, addition of the fluorine atom adjacent to the phenolic hydroxyl group reduces solubility by factors of ca. 2–10 for the matched pairs **4a** and **3g**, **4b** and **3h**, and **4c** and **3i**. This is the typical direction in view of the increase in hydrophobicity, except in special cases where the fluorine is 2 or 3 carbon atoms removed from an oxygen.^{5,17} Lengthening the linker by two methylene units in going from **4c** to **4d** reduced the solubility from 37 to 19 $\mu\text{g/mL}$; however, addition of an ethyleneoxy group in progressing to **4e** was very beneficial, yielding a 23-fold boost to 867 $\mu\text{g/mL}$. For **5a** and **5b**, the improvements with a carboxylic acid group are again apparent. Then, with **6**, removal of the intermolecular hydrogen bond in the crystal was tested; the result of 1046 $\mu\text{g/mL}$ reflects a 3-fold increase over the solubility of **3i**.

In Table 1, computed octanol/water partition coefficients, ClogP and QPlogP, have been included as determined by ChemDraw and QikProp.^{18,19} These are of interest as a measure of hydrophobicity and since log P and log S values are known to be correlated. The solubility equation of Yalkowsky estimates $\log S = 0.5 - \log P - 0.01(t_m - 25)$, where t_m is the melting point,^{2,5,20} and a simpler relationship is $\log S = -\log P - 0.2$ with

an rms error of 1.0 log unit.⁹ Thus, log S does decrease linearly with increasing hydrophobicity, as represented by log P, but much quantitative uncertainty is associated with the crystalline state, as reflected in t_m and other terms that have been introduced to represent it in predictive methods.² The log P predictions with the present methods are in generally good accord with an average difference of 0.62, though the discrepancies for **3c**, **3h**, and **4b** are greater than 1 log unit. Presently, log P is not found to be a good predictor of log S. For **3a–3e**, the log P values would suggest solubilities in the 10^{-3} to 10^{-4} M range, while the observed values are closer to 10^{-5} M, owing presumably to the π -stacking in the crystals (Figure 2). Differences in log P are also not a good gauge of differences in log S for the present compounds. For example, **5a** might be expected to be 100-fold less soluble than **3d** based on log P, but it is more soluble; similarly, the enhanced solubility of **4e** is not reflected in its log P.

Concerning the Yalkowsky equation, routine measurement of melting points is not common today in drug discovery settings. However, we did measure the melting points of several of the compounds. They are 200–205 °C for **3d**, 268–270 °C for **4d**, and 230–235 °C for **4e**. Application of the Yalkowsky equation using the average of the two log P values in Table 1 then yields predicted log S values of -5.0 , -5.9 , and -5.0 for the three compounds, respectively, which translate to 3.5, 0.5, and 3.9 $\mu\text{g/mL}$. Thus, compared to the experimental solubilities in Table 1, the prediction for **3d** is accurate, but the enhanced solubilities for **4d** and **4e** are completely missed. Even differences in solubility are difficult to predict, and, if solubility is an important issue, there is no alternative but to measure it.

Finally, for the K_i results, the tautomerase assay was repeated for all of the listed compounds in this report in order to limit variations from the protein preparation, incubation times, and spectrometer.¹³ For **3a–3i**, the activities are mostly in a narrow range of 0.2–0.4 μM , which is expected since the structural variations are only for the solvent-exposed substituent; additional modulation may come from interactions with residues on the surface of MIF. For example, the greater potency for **3e** may result from the $\text{CH}_3\text{O}(\text{CH}_2\text{CH}_2\text{O})_2$ appendage curling back and making hydrophobic contacts with the rings of Pro33 and Pro34, which are in the gray area above Lys32 in Figure 1. The present K_i results for **3a**, **3g**, and **3h** also agree well with K_d values that we reported recently for these compounds from a fluorescence polarization assay.¹⁴ As found previously,^{9,14} addition of the fluorine in going from **3** to **4**, e.g., **3g** to **4a** or **3h** to **4b**, enhances the activities ca. 3-fold, owing to contact with Met101 and enhancement of the hydrogen bond between the phenolic hydroxyl group and Asn97. Modeling also suggested introduction of the phenoxy group at C5 of the quinoline fragment to pick up contacts with Tyr36,^{9,14} which has yielded the most potent compounds, **5a** and **5b**. In addition, the importance of the hydrogen bond between the phenolic OH and Asn97 was confirmed by the increase in K_i to 1.37 μM for **6** from 0.20 μM for **3i**. As an additional control, the reference MIF inhibitor (R)-ISO-1 (**7**)²¹ was also assayed; the observed K_i of 27.3 μM is similar to the prior average value of 24 μM from multiple measurements.¹⁴

The principal purpose of this study was to explore systematically the effects of variations of a solvent-exposed substituent on

aqueous solubility in a drug-like series. The related work in the literature is largely scattered and less extensive, though the matched pair study of Zhang et al.⁴ and the study of polyether and alcohol substituents by Zhu et al.²² are particularly notable. The sequence of substituents in Table 1 may be of use to others who are faced with a similar challenge, though the quantitative outcomes will undoubtedly be different for other molecular series. However, the present and earlier results^{5,7} do point to the utility of the addition of polyether chains terminated with a morpholine ring or surrogates for improving aqueous solubility. In the present case, addition of polyether chains with a carboxylic acid was also highly effective.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsmmedchemlett.6b00451.

Synthetic procedures, NMR and HRMS spectral data for all new compounds, and crystallographic details for **3a** (PDF)

Crystallographic data for **3a** (CIF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: william.jorgensen@yale.edu.

ORCID

William L. Jorgensen: 0000-0002-3993-9520

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Gratitude is expressed to the National Institutes of Health (GM32136) for research support and to the National Science Foundation for a fellowship for M.J.R. (DGE-1122492).

■ ABBREVIATIONS

DCM, dichloromethane; THP, tetrahydropyran; TFA, trifluoroacetic acid

■ REFERENCES

- (1) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.
- (2) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54*, 355–366.
- (3) Stegemann, S.; Leveiller, F.; Franchi, D.; de Jong, H.; Lindén, H. When poor solubility becomes an issue: From early stage to proof of concept. *Eur. J. Pharm. Sci.* **2007**, *31*, 249–261.
- (4) Zhang, L.; Zhu, H.; Mathiowetz, A.; Gao, H. Deep understanding of structure-solubility relationship for a diverse set of organic compounds using matched molecular pairs. *Bioorg. Med. Chem.* **2011**, *19*, 5763–5770.
- (5) Walker, M. A. Improving Solubility by Structural Modification. *Top. Med. Chem.* **2013**, *9*, 69–106.
- (6) Jorgensen, W. L. Computer-aided discovery of anti-HIV agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768–4778.
- (7) Bollini, M.; Cisneros, J. A.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. Optimization of diarylazines as anti-HIV agents with dramatically enhanced solubility. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5213–5216.
- (8) Ishikawa, M.; Hashimoto, Y. Improvement in Aqueous Solubility in Small Molecule Drug Discovery Programs by Disruption of Molecular Planarity and Symmetry. *J. Med. Chem.* **2011**, *54*, 1539.
- (9) Dziedzic, P.; Cisneros, J. A.; Robertson, M. J.; Hare, A. A.; Danford, N. E.; Baxter, R. H.; Jorgensen, W. L. Design, Synthesis, and Protein Crystallography of Biaryltriazoles as Potent Tautomerase Inhibitors of Macrophage Migration Inhibitory Factor. *J. Am. Chem. Soc.* **2015**, *137*, 2996–3003.
- (10) Dahlgren, M. K.; Schyman, P.; Tirado-Rives, J.; Jorgensen, W. L. Characterization of Biaryl Torsional Energetics and its Treatment in OPLS All-Atom Force Fields. *J. Chem. Inf. Model.* **2013**, *53*, 1191–1199.
- (11) Crystals of **3a** were grown via vapor diffusion with DMSO and water. A resulting crystal was diffracted under a cold nitrogen stream on a Rigaku MicroMax 007HF+ equipped with a Saturn 944+ CCD detector with Cu K α . The structure was solved with direct methods. Heavy atoms were refined anisotropically, and hydrogen atoms were refined using a riding model. The hydrogen atom isotropic displacement parameters were fixed to the U values of the atoms to which they are linked, multiplied by 1.2. The molecules pack well with a pseudoslipped stacked morphology. The median plane formed by the triazole ring creates angles of 8.0(3)° and 23.9(4)° with the quinoline and phenol groups, respectively.
- (12) Baka, E.; Comer, J. E. A.; Takács-Novák, K. Study of equilibrium solubility measurement by saturation shake-flask method using hydrochlorothiazide as model compound. *J. Pharm. Biomed. Anal.* **2008**, *46*, 335–341.
- (13) Cisneros, J. A.; Robertson, M. J.; Valhondo, M.; Jorgensen, W. L. Irregularities in enzyme assays: The case of macrophage migration inhibitory factor. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 2764–2767.
- (14) Cisneros, J. A.; Robertson, M. J.; Valhondo, M.; Jorgensen, W. L. A Fluorescence Polarization Assay for Binding to Macrophage Migration Inhibitory Factor and Crystal Structures for Two Potent Inhibitors. *J. Am. Chem. Soc.* **2016**, *138*, 8630–8638.
- (15) Etter, M. C. Encoding and decoding hydrogen-bond patterns of organic compounds. *Acc. Chem. Res.* **1990**, *23*, 120–126.
- (16) The primary amines **3c**, **3g**, and **4a** were prepared as their TFA salts. This resulted from use of trityl protecting groups that are removed with TFA, yielding the salts. Attempted isolation of the free amines by treatment with NaHCO₃ and extraction with DCM or ethyl acetate gave poor yields, owing to low solubility of the amines.
- (17) Lee, W.-G.; Chan, A. H.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. Design, Conformation, and Crystallography of 2-Naphthyl Phenyl Ethers as Potent Anti-HIV Agents. *ACS Med. Chem. Lett.* **2016**, *7*, 0000–0000.
- (18) ChemBioDraw 12.0.2; CambridgeSoft Inc.: Cambridge, MA, 2014.
- (19) QikProp 4.0; Schrödinger Inc.: New York, NY, 2014.
- (20) Ran, Y.; He, Y.; Yang, G.; Johnson, J. L. H.; Laskowski, S. H. Estimation of aqueous solubility of organic compounds using the general solubility equation. *Chemosphere* **2002**, *48*, 487–509.
- (21) Lubetsky, J. B.; Dios, A.; Han, J.; Aljabari, B.; Ruzsicska, B.; Mitchell, R.; Lolis, E.; Al-Abed, Y. The tautomerase active site of macrophage migration inhibitory factor is a potential target for discovery of novel anti-inflammatory agents. *J. Biol. Chem.* **2002**, *277*, 24976–24982.
- (22) Zhu, G.-D.; Arendsen, D. L.; Gunawardana, I. W.; Boyd, S. A.; Stewart, A. O.; Fry, D. G.; Cool, B. L.; Kifle, L.; Schaefer, V.; Meuth, J.; Marsh, K. C.; Kempf-Grote, A. J.; Kilgannon, P.; Gallatin, W. M.; Okasinski, G. F. Selective Inhibition of ICAM-1 and E-Selectin Expression in Human Endothelial Cells. 2. Aryl Modifications of 4-(Aryloxy)thieno[2,3-c]pyridines with Fine-Tuning at C-2 Substituents. *J. Med. Chem.* **2001**, *44*, 3469–3487.