Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Discovery of Ligands for ADP-Ribosyltransferases via Docking-Based Virtual Screening

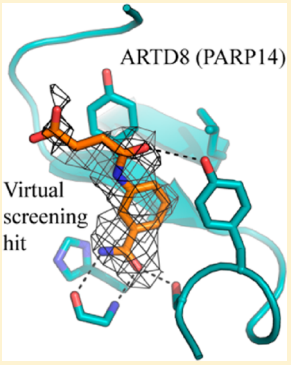
C. David Andersson,[†] Tobias Karlberg,[‡] Torun Ekblad,[‡] Anders E. G. Lindgren,[†] Ann-Gerd Thorsell,[‡] Sara Spjut,[†] Urszula Uciechowska,[†] Moritz S. Niemiec,[†] Pernilla Wittung-Stafshede,[†] Johan Weigelt,[‡,§] Mikael Elofsson,[†] Herwig Schüler,*[,‡] and Anna Linusson*[,†]

[†]Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden

[‡]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

Supporting Information

ABSTRACT: The diphtheria toxin-like ADP-ribosyltransferases (ARTDs) are an enzyme family that catalyzes the transfer of ADP-ribose units onto substrate proteins by using nicotinamide adenine dinucleotide (NAD^+) as a cosubstrate. They have a documented role in chromatin remodelling and DNA repair, and inhibitors of ARTD1 and 2 (PARP1 and 2) are currently in clinical trials for the treatment of cancer. The detailed function of most other ARTDs is still unknown. By using virtual screening, we identified small ligands of ARTD7 (PARP15/BAL3) and ARTD8 (PARP14/BAL2). Thermal-shift assays confirmed that 16 compounds, belonging to eight structural classes, bound to ARTD7/ARTD8. Affinity measurements with isothermal titration calorimetry for two isomers of the most promising hit compound confirmed binding in the low micromolar range to ARTD8. Crystal structures showed anchoring of the hits in the nicotinamide pocket. These results form a starting point in the development of chemical tools for the study of the role and function of ARTD7 and ARTD8.



INTRODUCTION

The diphtheria toxin-like ADP-ribosyltransferases (ARTDs) catalyze the transfer of ADP-ribose moieties onto either acceptor proteins or growing ADP-ribose chains, using nicotinamide adenine dinucleotide (NAD^+) as a cosubstrate. The human ARTD family comprises 18 members,[1] including the proteins formerly referred to as poly(ADP-ribose) polymerases (PARPs). ARTD1 (PARP1), the most abundant in human and best-characterized family member, is involved in DNA repair, chromatin modification, and cell death.[2] Inhibitors of the catalytic activity of ARTD1 are currently in clinical trials for the treatment of various cancer types.[3] Other family members include the B-aggressive lymphoma (BAL) proteins ARTD7 (PARP15/BAL3) and ARTD8 (PARP14/BAL2),[4] both of which are predicted mono(ADP-ribose) transferases.[5] Whereas ARTD7 remains poorly characterized, ARTD8 is implicated in Stat6-dependent transcription control[6] as well as in cytokine-regulated control of cellular metabolism.[7] Potent and selective small-molecule inhibitors are powerful chemical tools for the elucidation of cellular functions of individual enzyme family members.[8] Unfortunately, the majority of the commonly used PARP inhibitors affect several ARTD family members in vitro, and potent effectors of most ARTD enzymes are lacking altogether.[9]

The NAD^+ binding site of ARTDs has been identified from structural alignment of ARTDs with the crystal structure of diphtheria toxin in complex with NAD^+.[10] Crystal structures of ARTD catalytic domains in complex with ligands that mimic the nicotinamide moiety in NAD^+ showed similarity to diphtheria toxin,[10] and a common mode of ligand anchoring to conserved residues in their binding pockets.[9,11−15] The loop regions surrounding the NAD^+ binding site show a higher sequence and structural variability, in particular in the donor-site loop (D-loop), a critical element for both NAD^+ binding and substrate recognition.[16] The structural plasticity of the D-loop has been observed when comparing crystal structures in the ARTD family.[9,14] However, in many instances, the whole D-loop could not be modeled as a result of weak or absent electron density. In addition, in some crystal structures these parts of the complex participate in crystal packing, which makes a structural interpretation of ligand−D-loop interactions problematic.

In this study, we applied structure-based virtual screening using molecular docking to identify new and selective leads that can be developed into chemical tools for the study of the function of ARTD7 and ARTD8. Inhibitors of ARTD1 have previously been identified using ligand-based[17] or structure-based virtual screening methods,[18,19] showing that the NAD^+ binding site of the ARTDs is a viable screening target. Docking is a standard method for virtual screening, although it is well-known that the method has its limitations, for example, the under-sampling of protein conformations and the use of approximate scoring functions.[20] Nevertheless, structure-based virtual screening has proven its benefits in identifying new chemical leads.[21,22] The use of multiple scoring functions has

Received: May 29, 2012
Published: July 23, 2012

dx.doi.org/10.1021/jm300746d | J. Med. Chem. 2012, 55, 7706−7718

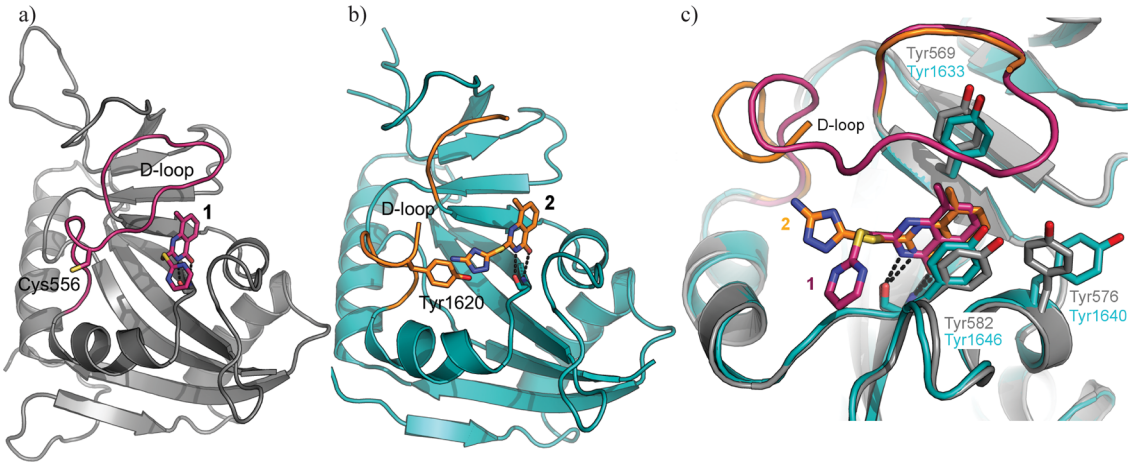


Figure 1. Inhibitors bound in the NAD$^+$ binding pocket display a common fork-motif interaction with a conserved glycine residue. Mimicking nicotinamide, they anchor using the amide group to backbone nitrogen and oxygen of this glycine. Overall structures of (a) ARTD7·**1** (purple) and (b) ARTD8·**2** (orange). (c) Overlay of binding sites in ARTD7·**1** (purple) and ARTD8·**2** (orange), with D-loops highlighted.

proved to be more robust when "optimal" scoring functions are unknown.[23−25] We have developed a virtual screening protocol targeting ARTDs 7 and 8 based on new protein structural information and a developed resemblance scoring approach. In short, a principal component analysis (PCA)-based evaluation that makes use of multiple docking poses and multiple scoring functions was used to identify hit molecules with similar scoring profiles to those of known ARTD inhibitors. The hit molecules from our virtual screen were compared with the hit molecules found when using a single scoring function (glidescore). The virtual screening hits were evaluated for binding to ARTD7, ARTD8, and ARTD1.

RESULTS AND DISCUSSION

Crystal Structures of ARTD7 and ARTD8: Structural Basis for Virtual Screening. Crystallization trials with ARTD7 and ARTD8 were performed in the presence of ligands that had previously been shown to bind to the proteins.[9] The complex of ARTD7 with the nonselective ligand **1** (STO1102) was solved and refined to a resolution of 2.4 Å (PDB ID: 4F0E). The complex of ARTD8 with the nonselective ligand **2** (STO1190) has been presented previously (PDB ID: 3SMI; resolution 2.4 Å).[9] Overall, both structures have a similar fold to previously reported structures of ARTDs but they lack the amino-terminal α-helical subdomain that is present in ARTD1−4 (Figure 1).[11−13,26] A superposition of ARTD7·**1** and ARTD8·**2** revealed a high structural similarity in the NAD$^+$ binding site (Figure 1), where the 8-methylquinazolin-4-(3*H*)-one moiety of **1** is stacked between two tyrosines (Tyr569 and Tyr582) and forms the conserved hydrogen-bond interactions with Gly538 situated at the end of β-sheet **3**. A conformational difference is seen for Tyr576, as this residue perturbs the NAD$^+$ binding site significantly more than does the corresponding Tyr1640 of ARTD8. A notable sequence difference in the D-loops, with effects on ligand binding, is in the position of ARTD8-Tyr1620, corresponding to ARTD7-Cys556. The triazol-amine of **2** stacks with the Tyr1620 side-chain, an interaction that is absent in the ARTD7·**1** complex.

A comparison of previously determined crystal structures of ARTD7, ARTD8, and other ARTD catalytic domains in complex with nicotinamide mimics reveals structural differences in the D-loop (structures listed in the Supporting Information).[9,14] Whereas the D-loop is fully intact in the ARTD7·**1** complex, we did not observe any electron density for ARTD8 D-loop residues 1622−1626, indicating that the loop is flexible. For the purpose of virtual screening, we therefore used multiple protein structures, including two different structural models each for ARTD7 and ARTD8: one with the coordinates as determined by crystallography (referred to as "long"), and a truncated version in which parts of the D-loop were omitted (referred to as "short"; see the Experimental Section for details).

Virtual Screening with ARTD7 and ARTD8. The virtual screening procedure is presented in Figure 2. A virtual library of 8050 molecules, including known ligands (12477 tautomers and different protonation states), were docked to two versions each of ARTD7 and ARTD8, and a maximum of 100 docked poses per molecule were saved (Figure 2A,B). Molecules that did not exhibit the conserved hydrogen bond interactions with Gly on β-sheet **3** were discarded, and 2904, 4731, 6476, and 6805 molecules, respectively, remained as ligands of ARTD7$_{long}$, ARTD7$_{short}$, ARTD8$_{long}$, and ARTD8$_{short}$, respectively (Figure 2C). This corresponded to 26000−120000 docked poses, depending on the protein structure. The docked poses were subsequently rescored with eight scoring functions, and the score values were normalized with respect to the number of heavy atoms in the docked molecules (Figure 2D,E).

The univariate statistics of the multiple poses (Figure 2F) for each of the five and six, respectively, known ligands of ARTD7 and ARTD8[9] (see the Supporting Information for structures) revealed that all but one (the ARTD7·2-dimethylamino-*N*-(6-oxo-5,6-dihydro-phenanthridin-2-yl)-acetamide (STO802) complex) of the known ligands formed the conserved interactions with Gly for at least six, but most often, for more than 80 poses. Comparisons of the scoring values for the known ligands and the virtual library molecules showed that the known binders in general were scored to be good binders. In addition, comparison of the docked poses with the crystallographically determined poses (**1** in ARTD7 and **2** in ARTD8) confirmed the reproduction of the bioactive conformations of at least one pose to a root-mean-square deviation (rmsd) of at most 0.9 Å.

The multivariate analysis of the docking score data (i.e., the docking score resemblance approach) for each of the four

dx.doi.org/10.1021/jm300746d | J. Med. Chem. 2012, 55, 7706−7718

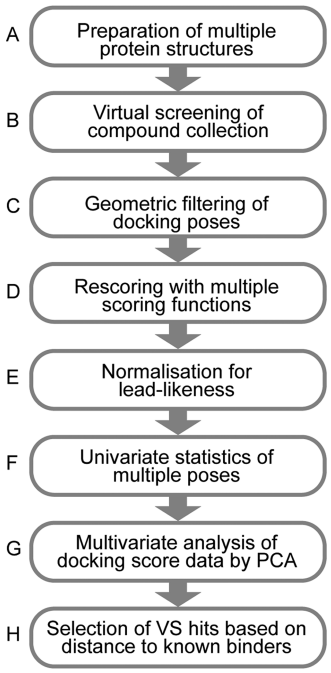


Figure 2. Flowchart of the virtual screening procedure (A–H), including the docking score resemblance approach (F–H). (A) Preparation of $ARTD7_{long}$, $ARTD7_{short}$, $ARTD8_{long}$, and $ARTD8_{short}$, differing in the D-loop amino acid composition. (B) Virtual screening of 12,477 molecules and known ligands. (C) Filtering to exclude molecules without two hydrogen bonds to the conserved Gly in the NAD+ binding site. (D) Scoring of docked poses with glidescore, goldscore, chemgauss3, chemscore, oechemscore, plp, zapbind, and screenscore. (E) Normalization of score values with respect to molecule size. (F) Scoring of each molecule, by weighing together scores for molecule-specific docking poses. (G) Compression and visualization of scoring data. (H) Selection of molecules which were similar, in terms of Euclidean distance in principal component space, to the known binders.

protein structures gave five principal components (PCs) for each of the four PCA models (Figure 2G; see the Supporting Information for model statistics). The patterns and interpretations were similar for ARTD7 and ARTD8; in the first two PCs, there was a clear clustering of the highly ranked known binders, as illustrated in Figure 3a, owing to their generally numerically low docking scores, low variation of scores across different poses, and high number of poses passing through the geometric filter. The pattern seen in the *t*-scores plot can be explained by the contributions made to each PC by the univariate variables (Figure 3b). PCs 3–5 described the influence of individual scoring functions on different molecules (for plots see the Supporting Information). Thus, the use of all five PCs to select virtual screening hits enabled us to take both "consensus scoring" and individual scoring functions into account at the same time. The 2904–6805 docked molecules showed a widespread in all five PCs, and the distance in PC space of a molecule to the centroid of the known binders was used to select hits, this distance is here called docking score resemblance (Figure 2H). The four PCA models based on scoring data from the virtual screens with $ARTD7_{long}$, $ARTD7_{short}$, $ARTD8_{long}$, and $ARTD8_{short}$ were considered and selection of all molecules located near or in the *t*-score volume occupied by the known binders in each of the models resulted in 64 molecules (set A). A second set of molecules (set B) was selected from the 64 molecules ranked highest in terms of the normalized glidescore values for all four protein structures. Forty-seven of these 64 molecules were not in set A. Set A and set B were biologically evaluated for binding to ARTD7 and ARTD8.

Evaluation of Virtual Screening Hits. The molecules that experimentally displayed binding to ARTD7 and ARTD8 are presented in Table 1. Out of the 111 screened compounds (64 in set A and 47 in set B), 16 stabilized ARTD7 or ARTD8 according to differential scanning fluorimetry (cutoff, $\Delta T_m \geq 1.5$ °C). The purity of the hits was confirmed to be >95% according to reversed-phase HPLC UV trace, and the structures were confirmed using ^{1}H NMR spectroscopy (see spectra in the Supporting Information). Surprisingly, the spectrum for the pure compound **A16** revealed a (*Z*)-configuration, compound **6** ($A16_{(Z)}$), as opposed to the (*E*)-configuration, compound **5** ($A16_{(E)}$), reported in the database, and the ΔT_m value reported in Table 1 is thus for the (*Z*)-isomer **6**. The majority of hits bound to ARTD8 (14 molecules), while two bound to ARTD7. The largest ΔT_m (>3 °C) was detected for binders to ARTD8. The imbalance in the numbers of binders to the two targets was seen throughout the screening; more docking poses in total "survived" the dockings to ARTD8, and a higher number of molecules targeting ARTD8 were selected than for ARTD7 (see the Supporting Information). A possible explanation for these observations is that the ARTD7 NAD^+ binding site was partially constricted by the unique positioning of Tyr576. In contrast, the ARTD8 NAD^+ binding site was somewhat larger and less constricted by the D-loop, allowing for more molecules to adopt a probable binding mode in the screening. Notably, the two compounds that bound to ARTD7 and five of the compounds that bound to ARTD8 were selected from the virtual screens targeting ARTD8 and ARTD7, respectively. Some overlap of the results was expected, as the two proteins have very similar binding sites when alternative D-loop conformations are neglected and because docking is a rather crude method that mainly considers shape complementarities and key electrostatic interactions such as hydrogen bonds. Examples of docking poses for eight structurally representative compounds are given in the Supporting Information. The poses displayed by the compounds selected by our docking score resemblance approach differed compared to the X-ray crystal pose of ligand **2**.

Importantly, no single screen against one protein structure identified all hits. In general, more hits (63%) were identified in the screen against ARTDs with unaltered D-loops, and we reason that the volume of the binding site in these structures may be closer to the "true" volume. Most hits (38%) were selected from the screen against $ARTD8_{long}$. Notably, three hits (**5**, **9**, and **13**) were exclusively identified from the dockings against ARTDs with truncated loops. These observations support the theory that there is a benefit in using several structures for the same protein and/or related proteins to increase the chance of finding molecules that bind. It is notable that the two scoring methods, docking score resemblance and glidescore, identified nine ligands each with only two ligands being identified by both methods, suggesting that the two methods are complementary.

Analysis of Virtual Screening Hits. It has been reported that the property space for chemical probes is generally closer to that of chemical leads than that of drugs,[33] so we analyzed our hits for lead-likeness. The majority of our identified binders showed lead-like properties, with a molecular weight (M_r) of 100–350 and calculated LogP (CLogP) of 1.0–3.0.[34] The hits

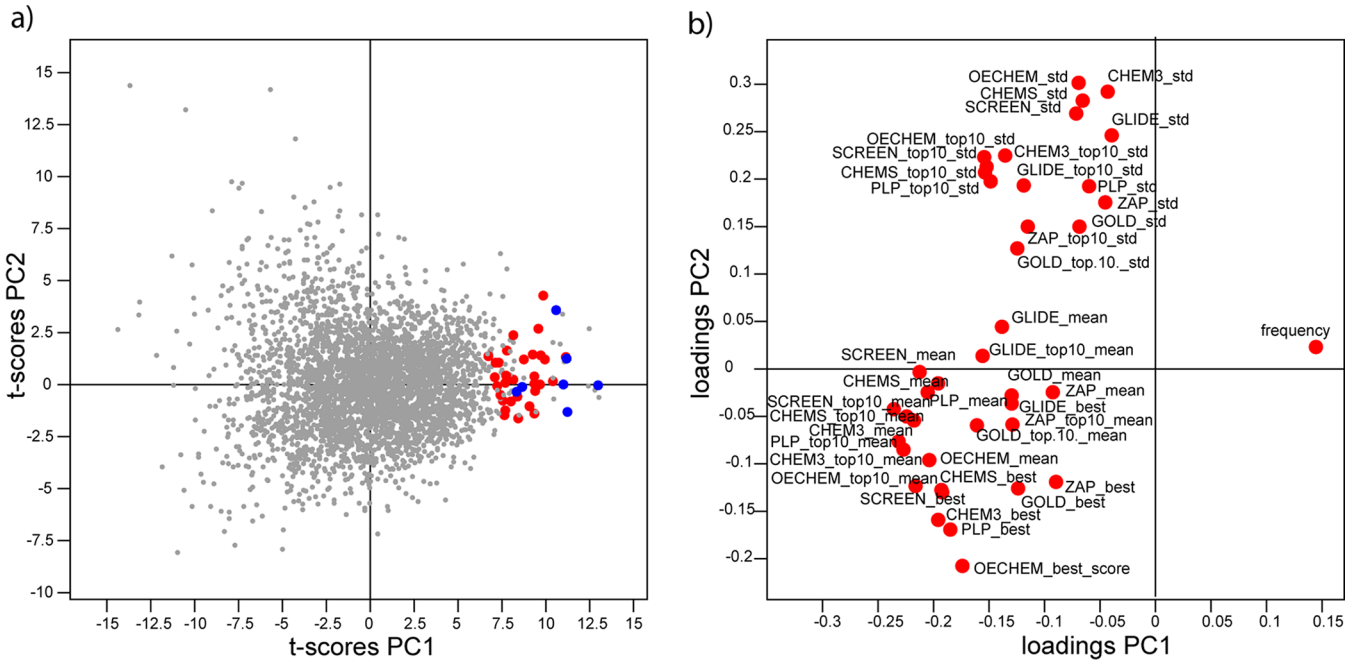


Figure 3. PCA *t*-scores and loadings plots for ARTD8$_{long}$, illustrating the virtual screening results and the selection of molecules based on docking score resemblance. (a) Plot of PCA *t*-scores of PC1 vs PC2, where all molecules from the virtual library, known binders, and selected molecules are indicated in gray, blue, and red, respectively. Note that the selected molecules were close to the known binders with respect to all five PCs, not only the two shown here. (b) Loadings plot illustrating how the various scoring functions influence the molecules' positions in the *t*-scores plot. Abbreviations are as follows: GLIDE, glidescore; CHEM3, chemgauss3; CHEMS, chemscore; OECHEM, oechemscore; SCREEN, screenscore; GOLD, goldscore; ZAP, zapbind; std, standard deviation of mean. Frequency is the number of poses after filtration.

had an M_r within this span and a CLogP within the span or slightly lower (**19** had a CLogP of −1.2). The exception for lead-likeness was **7**, which had a higher M_r of 359 and a higher CLogP of 4.6.

A structural diversity analysis of the 16 hits based on MACCS fingerprint descriptors[35] and subsequent clustering resulted in eight clusters which are presented in Figure 4. The hit molecules identified by docking score resemblance (set A) belonged to six different clusters, while the hits identified by glidescore (set B and overlapping molecules **11** and **12**) belonged to three different clusters. It is clear that docking score resemblance, compared to glidescore, gave a more diverse set of hit molecules with respect to their structural features. Clusters 6 and 8, which contain five and two hits, respectively, were identified by glidescore, whereas compounds in cluster 1 were identified with both methods. Five of the clusters contained lone compounds, all identified by our docking score resemblance approach. It has been shown that even quite similarly operating scoring functions favor different ligands as top candidates in virtual screening,[36,37] which would imply that considering several scoring functions could also increase the diversity among the top candidates, as seen in our results.

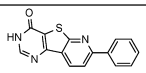
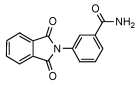
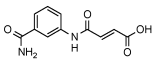
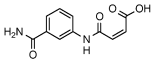
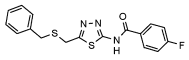
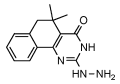
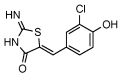
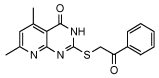
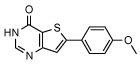
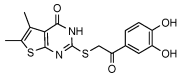
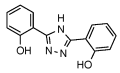
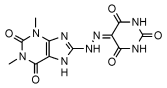
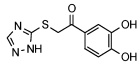
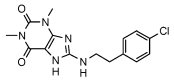
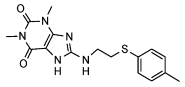
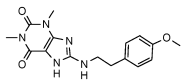
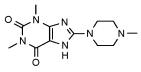
ARTD Binding Profile. To investigate the binding profile and selectivity potential of our hit molecules, we complemented the binding data of ARTD7 and ARTD8 with ΔT_m data for ARTD1 (Figure 5). All but one of the hits induced a greater thermal stability shift in ARTD7 or ARTD8 than in ARTD1. This is promising because all previously reported binders to ARTD7 and ARTD8 possess a stronger binding to ARTD1.[9] Here, we defined hit molecules to have selectivity potential if the hit had a $\Delta T_m \geq 3$ °C and the ΔT_m was 1.5 °C higher for one ARTD than for the other two. Four compounds (**5**, **6**, **9**, and **10**)[38] belonging to three different compound classes (Figure 4), fulfilled the criteria and bound to ARTD8 with a potential selectivity over ARTD1 (Figure 5). None of the hits had potential selectivity for ARTD7. Note that seven compounds among the 111 tested only bound to ARTD1 (see structures in the Supporting Information) and that compounds **22** and **25** had potential selectivity for ARTD1 over ARTD7/ARTD8.

The compound classes were investigated for their suitability as candidates for hit-to-lead (chemical tool) transformation, in terms of ΔT_m, potential selectivity, structural features, and synthetic feasibility. Cluster 1 contained the binders for ARTD7 (**3** and **11**) but also molecules binding to ARTD1 (**12**) and ARTD8 (**10**), which may indicate a selectivity issue with this class of compounds. We note that among the binders to ARTD1 there are clear structural similarities and that the position of the amide in the heteroaromatic ring system is the same in these molecules, which may be an unwanted structural feature in selective molecules. Compound **9** in cluster 7, although displaying a high ΔT_m and selectivity toward ARTD8, was excluded from further investigations in this study. Cluster 5 (**8**) was discarded owing to its relatively low ΔT_m and poor selectivity. Cluster 6 contained five 8-substituted theophylline analogues, which were not studied further in this work. The remaining hit classes, clusters 2, 3, 4, and 8, were all considered to be of interest and pursued further. Compound **5** (in cluster 3), which bound with a high ΔT_m and showed potential selectivity for ARTD8 over ARTD1, was chosen for further investigation of its thermodynamic parameters and binding modes using isothermal titration calorimetry (ITC) and X-ray crystallography, respectively.

Investigation of Compounds 5 and 6 with ITC, X-ray Crystallography, and Molecular Dynamics Simulations. *Synthesis of Stereoisomers 5 and 6.* The virtual screening

dx.doi.org/10.1021/jm300746d | J. Med. Chem. 2012, 55, 7706−7718

Table 1. Thermal Shifts in Differential Scanning Fluorimetry (ΔT_m) Induced by Molecules from the Virtual Screening Selected According to Docking Score Resemblance (set A) and/or glidescore (set B)

			ARTD7		ARTD8		
	Name	Compound	mean ΔT_m	st. dev.	mean ΔT_m	st. dev.	Selection Criterion
3	A10[27]		**1.51**	0.37	0.65	0.84	resemblance
4	A14[28]		0.06	0.10	**2.10**	0.12	resemblance
5	$A16_{(E)}$		n.d.[a]	-	n.d.[a]	-	resemblance
6	$A16_{(Z)}$		0.80	0.10	**3.35**	0.67	-
7	A25		0.36	0.41	**1.66**	0.83	resemblance
8	A28		1.38	0.04	**2.05**	0.60	resemblance
9	A36		-0.25	0.12	**3.89**	0.91	resemblance
10	A39		-0.09	0.09	**2.65**	0.12	resemblance
11	AB11[29]		**1.49**	0.10	1.11	0.62	resemblance / glidescore
12	AB40[30]		0.33	0.27	**1.63**	1.05	resemblance / glidescore
13	B11[31]		-0.34	0.07	**2.65**	1.84	glidescore
14	B18		-0.08	0.03	**2.12**	0.00	glidescore
15	B21		-0.05	0.14	**2.28**	0.12	glidescore
16	B39		0.66	0.07	**3.01**	0.67	glidescore
17	B40		0.09	0.13	**2.11**	0.69	glidescore
18	B41		1.09	0.04	**2.62**	0.02	glidescore
19	B42[32]		0.19	0.11	**1.56**	0.06	glidescore

[a]Not determined.

identified **5** with the (*E*)-configuration, and we decided to synthesize the (*E*)- and (*Z*)-configurations in order to determine the binding affinities of both compounds. The (*E*)- and (*Z*)-isomers were synthesized in two and one steps, respectively, to give pure **5** and **6** (Schemes 1 and 2, respectively).

*Measurement of Binding Affinities of the Isomers **5** and **6**.* To further characterize the binding of the (*E*)-isomer **5** and the (*Z*)-isomer **6** to ARTD8, we determined their binding properties using ITC. The ITC measurements of **5** and **6** show that both isomers bind to ARTD8, with similar dissociation constants in the low micromolar range (Figure 6). The K_d values (K^{-1}) for **5** and **6** were determined to be 11.2 and 7.6 μM, respectively. The binding event was mainly enthalpy driven in both cases; for **5** and **6**, the ΔH was −7.7 and −7.1 kcal/mol, respectively, while the $-T\Delta S$ was 0.92 and 0.10 kcal/mol, respectively. The magnitude of the K_d values measured for the two isomers indicate that the ΔT_m values from the thermal stability assay can be used as a simple way to

dx.doi.org/10.1021/jm300746d | J. Med. Chem. 2012, 55, 7706−7718

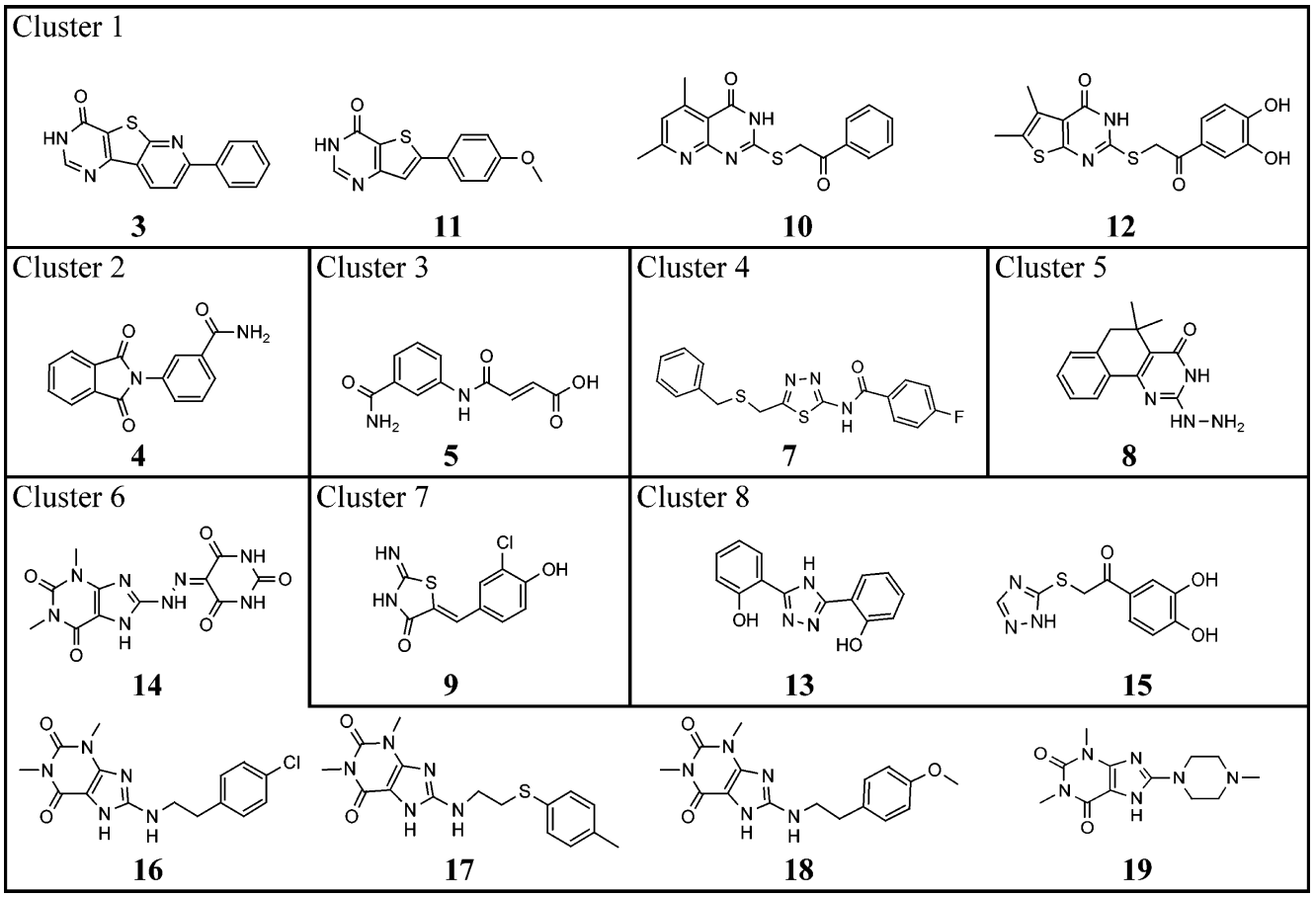


Figure 4. Tanimoto clustering of the virtual screening hits based on MACCS fingerprints; molecules with similar or common structural fragments are clustered together.

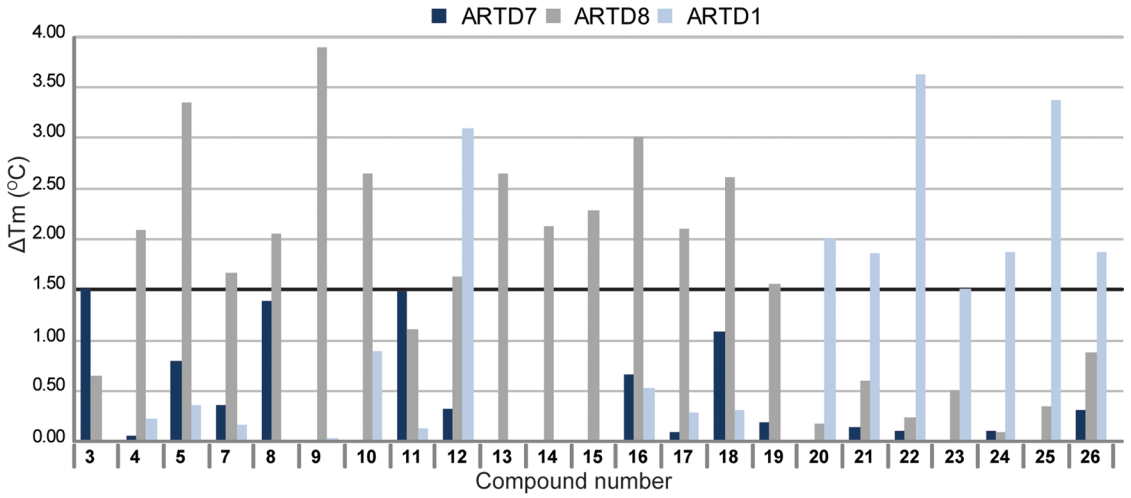


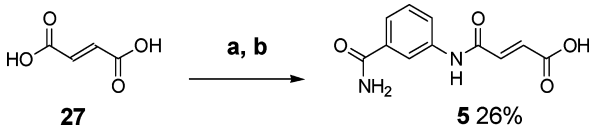
Figure 5. Compounds selected by docking score resemblance and/or glidescore that had a $\Delta Tm \geq 1.5$ °C in ARTD7, ARTD8 or ARTD1.

find molecules that bind to the ARTDs in a desirable concentration range.

*Crystal Structures of ARTD8 in Complex with Compounds **5** and **6**: Structural Analysis, Molecular Dynamics Simulations, and Comparison to ARTD1.* The (*Z*)-isomer **6** from the compound collection was successfully cocrystallized with ARTD8 and the crystal diffracted to a resolution of 1.9 Å (PDB ID: 4F1L). The crystallographic unit cell contains four monomers (A–D), where A and C make crystal-packing contacts with B and D, respectively, which completely overlap with the NAD^+ binding sites (see the Supporting Information). This has previously been observed for other crystals of ARTD8·ligand complexes.[9] Compound **6** could unambiguously be modeled with a well-defined electron density in two monomer-specific conformations: conformer 1 (in monomer A and C) and conformer 2 (in monomer B and D, see Figure 7). For both conformers, there is an internal hydrogen bond between the carboxylic acid and the anilinic nitrogen, forming a

dx.doi.org/10.1021/jm300746d | *J. Med. Chem.* 2012, 55, 7706–7718

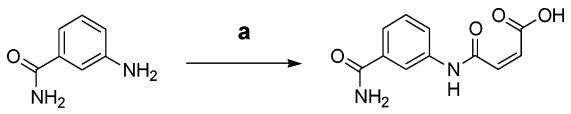
Scheme 1. Synthesis of (*E*)-4-(3-Carbamoylphenylamino)-4-oxobut-2-enoic Acid (5)[a]



27 **5** 26%

[a]Reagents and conditions: (a) $(COCl)_2$, dimethylformamide, CH_2Cl_2, 40 °C, 7 h; (b) 3-aminobenzamide, 1,4-dioxane, 40 °C, 16 h.

Scheme 2. Synthesis of (*Z*)-4-(3-Carbamoylphenylamino)-4-oxobut-2-enoic Acid (6)[a]



28 **6** 83%

[a]Reagents and conditions: (a) maleic anhydride, tetrahydrofuran, 18 h.

six-membered ring including the double bond. In terms of binding mode of **6** to ARTD8, both conformers bind with the benzamide moiety in the nicotinamide subsite of the NAD^+ binding pocket, where it is stacked between Tyr1633 and Tyr1646. The primary amide forms the typical hydrogen-bonding fork-motif interaction with Gly1602, as predicted by the docking, and has an additional interaction with the hydroxyl group of Ser1641. The (*Z*)-carbamoyl-4-oxobut-2-enoic acid moiety of **6** shows different binding modes in the two different conformers. Ligand conformer 1 (Figure 7a) makes interactions with residues of the D-loop of ARTD8: the amide carbonyl forms a hydrogen bond with the side-chain of Asn1624 and the position of the carboxylic acid hydrogen bonding to Tyr1640. Ligand conformer 2 (Figure 7b) displays a 180° rotation around the C−N bond by attaching the (*Z*)-carbamoyl-4-oxobut-2-enoic acid part to the phenyl. In monomer D, the amide carbonyl forms a hydrogen bond with the hydroxyl group of Tyr1640, while this interaction is absent in monomer B. The exact position of the D-loop could not be determined in this monomer owing to poor electron density, although it is clear that it is in an open conformation pointing away from the NAD+ binding site to avoid steric clashes of neighboring monomers.

The (*E*)-isomer **5** was cocrystallized with ARTD8 and the crystal diffracted to a resolution of 2.8 Å (PDB ID: 4F1Q), and the ligand binding site is shown in Figure 8. ARTD8·**5** formed a similar crystal lattice as for ARTD8·**6**, with crystal-packing contacts in the ligand binding pocket (see the Supporting Information), although the minimal asymmetric unit included two monomers (A and B). The D-loops appear to be in open or closed conformations similar to ARTD8·**6**, although it was not possible to model the complete D-loop for any monomers in this complex.

The two ligands in the monomers in the ARTD8·**5** complex were in similar conformations. The benzamide moiety of **5** bound in a comparable manner as for **6** (Figure 9a). The (*E*)-carbamoyl-4-oxobut-2-enoic acid moiety showed a binding mode similar to that of conformer 2 of compound **6** in

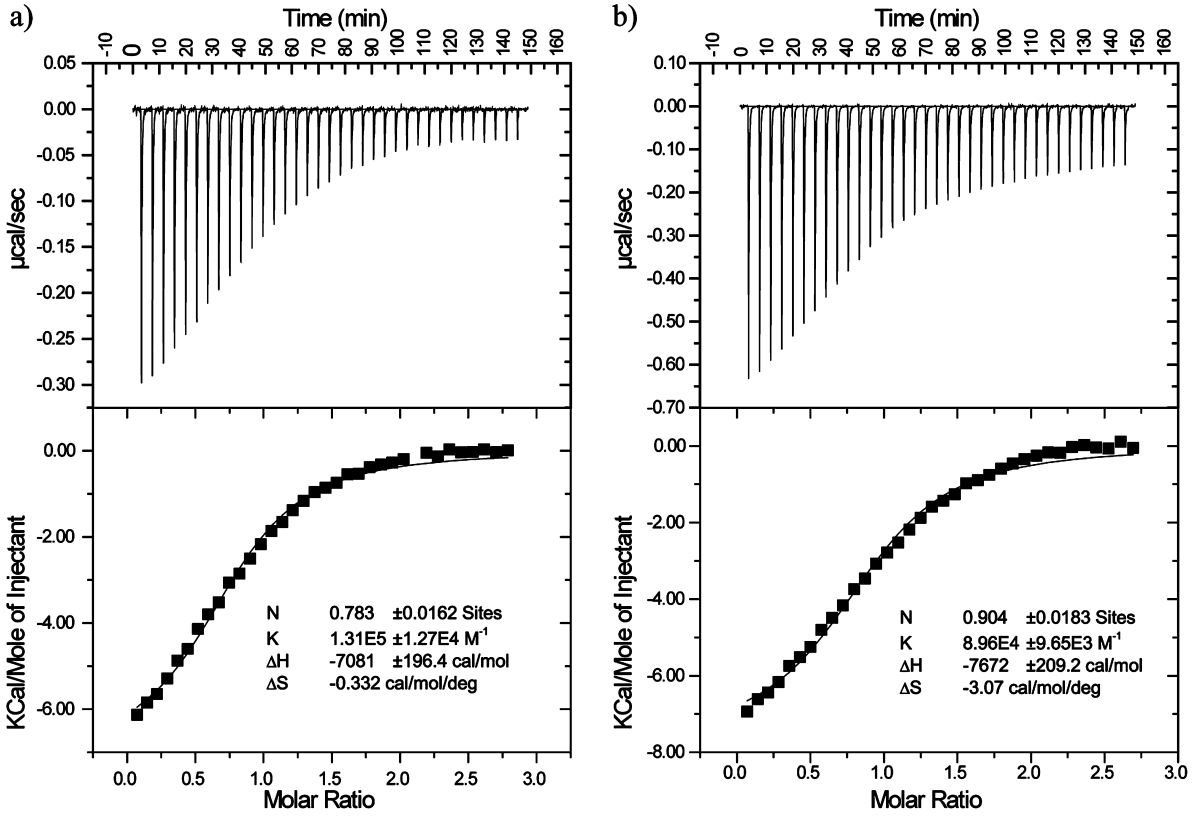


Figure 6. ITC results showing raw data after integration baseline correction (upper pane) and integrated data and regression (lower pane). *N* is the number of molecules per binding site, *K* is the association constant, ΔH is the change in enthalpy, and ΔS is the change in entropy. Error estimates are calculated from the regression fit of a single experiment. (a) ITC results for (*Z*)-isomer **6** and ARTD8. (b) ITC results for (*E*)-isomer **5** and ARTD8. The ITC experiments were performed at 25.0 °C.

dx.doi.org/10.1021/jm300746d | *J. Med. Chem.* 2012, 55, 7706−7718

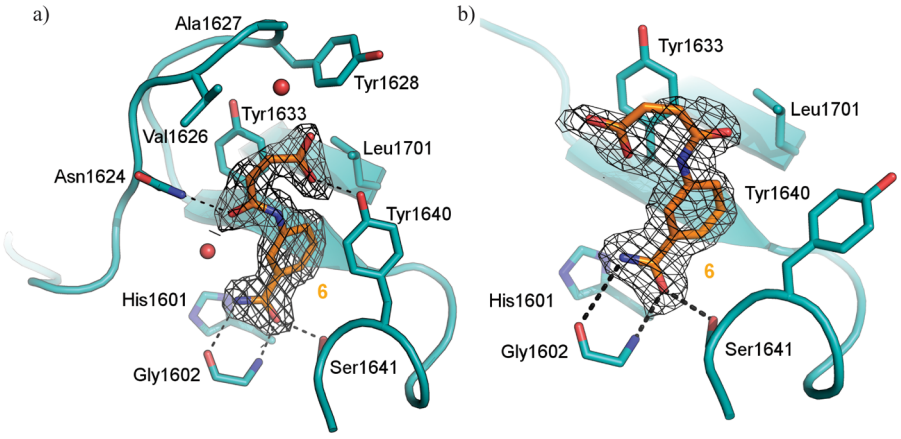


Figure 7. The two binding modes of (*Z*)-isomer **6** (orange sticks) bound to ARTD8, ($2F_{obs} - F_{calc}$) electron density map contoured at 1.0 σ (black wireframe). (a) ARTD8 monomer A with compound **6** conformer 1, and (b) ARTD8 monomer B with compound **6** conformer 2.

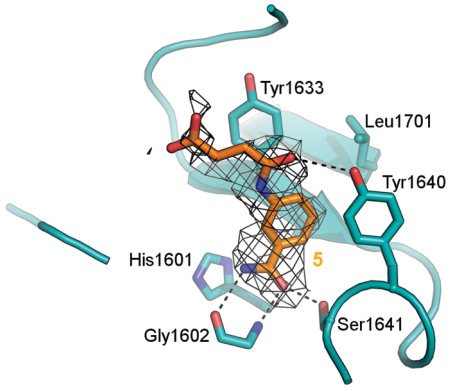


Figure 8. The binding mode of (*E*)-isomer **5** (orange sticks) with ARTD8 monomer A (cyan) and ($2F_{obs} - F_{calc}$) electron density map contoured at 1.0 σ (black wireframe).

monomer B, where the amide carbonyl forms a hydrogen bond (2.45 Å) with the hydroxyl group of Tyr1640.

The crystallographic data unambiguously showed that the (*E*)-isomer **5** and the (*Z*)-isomer **6** bind to the NAD^+ binding site in ARTD8. To select a monomer suitable as a basis for structure-based design, we investigated the implications of the crystal packing on single-monomer stability in the ARTD8·**6** and ARTD8·**5** complexes using molecular dynamics (MD) simulations. The calculations were performed in explicit water with modeled D-loops.

The MD simulation results showed that the noncovalent interactions in general and the hydrogen bond fork interaction with Gly1602 in particular were best preserved in the complex between ARTD8 (monomer B) and (*Z*)-isomer **6** (conformer 2) throughout 5 ns simulations. The average distances between **6** primary benzamide O and N to Gly1602 N and O was 3.0 and 3.5 Å, respectively, indicating a preserved hydrogen bond interaction (see the Supporting Information for distance plots). In contrast, the MD simulations of monomers C and D resulted in a loss of at least one of the hydrogen bond interactions between **6** and Gly1602 (distances >6 Å). The MD simulations for monomer A of ARTD8·**6** (conformer 1) resulted in a complete displacement of the ligand from the binding pocket after 300 ps, and the MD simulation was aborted after 600 ps. In the ARTD8·**5** crystal structure, monomer A displayed preserved interactions throughout the 5 ns MD simulation. The average distances between the primary benzamide O and N of compound **5** to Gly1602 N and O were 3.4 Å for both, while the MD simulations of monomer B resulted in a loss of the hydrogen bond interactions between **5** and Gly1602 (distances >4 Å). Comparing the two isomers, the conformer of **5** in monomer A and conformer 2 of compound **6** in monomer B

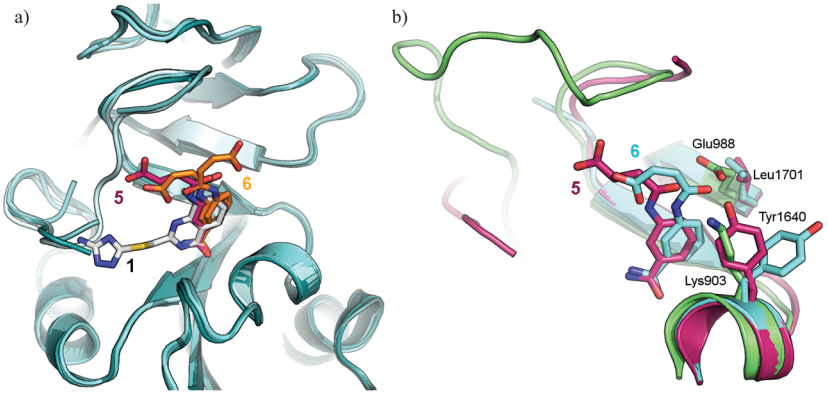


Figure 9. Overlay of (a) compound **2** (white sticks) from the ARTD8·**2** complex (PDB ID: 3SMI), ARTD8·**6** ligand conformer 1 and conformer 2 (orange sticks), with the acrylic acid pointing to the right and left, respectively, and the ARTD8·**5** ligand conformer (purple sticks) and (b) ARTD1 (PDB ID: 3L3M; green) and compound **6** conformer 2 in ARTD8 (PDB ID: 4F1L) monomer B (cyan) and compound **5** in ARTD8 (PDB ID: 4F1Q) monomer A (purple).

dx.doi.org/10.1021/jm300746d | *J. Med. Chem.* 2012, 55, 7706−7718

displayed similar interaction patterns with ARTD8 (e.g., hydrogen bonds to Gly1602 and Tyr1646) after 5 ns simulation, in addition to their similar binding strength measured by ITC. On the basis of the MD simulations, we argue that the ARTD8·**6** monomer B with ligand conformer 2 and ARTD8·**5** monomer A is anticipated in solution (i.e., is less affected by the crystal packing) and are the most valid starting points for structure-based design of chemical tools targeting ARTD8.

Compared to currently known ARTD ligands, the carbamoyl-4-oxobut-2-enoic acid substituent of the (*Z*)-isomer **6** (both conformers) and the (*E*)-isomer **5** point to a different part of the binding pocket of ARTD8 (Figure 9).[9] This part of the binding site corresponds to the less-conserved area among the ARTDs and shows variability in both structure and molecular properties. For example, in ARTD1, Tyr1640 is replaced by a lysine (Lys903) and Leu1701 is replaced by a glutamate (Glu988), both residues point toward the ligand binding site. Altogether, the differences in the binding sites result in a much more crowded part of the binding site for ARTD1 than for ARTD8. These observations are in agreement with the observed lower thermal stabilization of **5** and **6** for ARTD1 than for ARTD8.[38] The fact that the positions of the carbamoyl-4-oxobut-2-enoic acid moiety in **5** were correctly predicted by the dockings, and validated with the MD simulations, further strengthens the quality of the hit from a structure-based design point-of-view. The finding that the carbamoyl-4-oxobut-2-enoic acid in compounds **5** and **6** have a binding mode different from corresponding parts of ligands that bind to ARTD1 increases its potential as a chemical lead with selectivity for ARTD8 over ARTD1.

CONCLUSIONS

A set of 16 compounds was discovered that bind to the ADP-ribosyltransferases ARTD7 and/or ARTD8. The molecules belonged to eight different structural clusters, and three of the molecules displayed a potential selectivity for ARTD8 over ARTD1. We regard these classes of molecules as potential starting points in the development of selective chemical tools for elucidation of the role of ARTDs in cellular functions.

To identify the chemical starting points, multiple crystal structures of ARTD7 and ARTD8 were used in a structure-based virtual screening campaign aimed at identifying molecules that would bind to the NAD^+ binding sites in these ARTDs. The results showed that no single screen against one of these protein structures identified all hits, highlighting the benefit of using several different structures for the proteins in virtual screening. The screening protocol that was developed included molecular docking and a developed docking score resemblance approach. This approach was used to rank the molecules, i.e., to estimate the likelihood that a molecule would bind, and was based on similarity of scoring profiles of screened molecules to those of known inhibitors. It was evident that our scoring approach gave more structurally diverse hit molecules than did scoring with a single scoring function (glidescore), although the number of hits was the same for the two approaches.

Finally, we investigated the binding affinity and binding mode of the promising chemical leads **5** and **6**, which selectively bound to ARTD8. Both the isomers displayed an affinity for ARTD8 in the low micromolar range, suggesting that thermal shift was a valid indicator of affinity in this case. Both isomers bound to the NAD^+ binding site of ARTD8, according to the crystal structures. MD simulations of single monomers, based on the herein-determined crystal structures of ARTD8 in complex with compounds **6** and **5**, respectively, resulted in structures with maintained interaction patterns. The two isomers displayed a similar binding mode that is different from previously reported ligands in complexes with ARTD8 or ARTD1, which strengthen their potential as chemical leads with selectivity for ARTD8. Compounds **5** and **6** are small (M_r = 233) and display good water solubility.[39] They are thus good starting points for further optimization with the goal of fine-tuning the affinity and selectivity for ARTD8, in the search for new chemical tools for the ARTD protein family.

EXPERIMENTAL SECTION

Protein Crystallization. The recombinant production of human ARTD7 and ARTD8 is detailed in the Supporting Information. Crystals of the ARTD7·**1** complex (PDB ID: 4F0E) were obtained by the sitting-drop vapor-diffusion method in a 96-well plate (Corning) by mixing 0.1 μL of protein at a concentration of 31.3 mg/mL with 2 mM **1** dissolved in dimethyl sulfoxide (DMSO) and 0.1 μL of reservoir solution containing 40% 2-methyl-1,3-propanediol and 0.1 M *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) pH 10.5. The plate was incubated at 20 °C, and crystals appeared within two weeks. Crystals were frozen in liquid nitrogen, with no additional cryoprotectants.

Crystals of the ARTD8·**6** complex (PDB ID: 4F1L) were obtained by the sitting-drop vapor-diffusion method in a 96-well plate (Corning). Compound **6** was added in portions to a final concentration of 1.3 mM, to a protein solution (8.5 mg/mL), and the sample was concentrated 5-fold. Then 0.1 μL of the resulting protein−ligand solution was mixed with 0.2 μL of reservoir solution consisting of 20% (w/v) polyethylene glycol (PEG) 3350, 0.2 M sodium nitrate, and 0.1 M Bis-Tris pH 5.9. The plate was incubated at 20 °C, and crystals appeared within one month. The crystals were quickly transferred to a cryosolution consisting of 20% (w/v) PEG 3350, 0.2 M sodium nitrate, 0.1 M Bis-Tris pH 6.0, 20% (v/v) glycerol, and 0.2 M sodium chloride and then flash frozen in liquid nitrogen.

Crystals of the ARTD8·**5** complex (PDB ID: 4F1Q) were obtained by the sitting-drop vapor-diffusion method in a 96-well plate (Corning). Compound **5** was added incrementally to a final concentration of 1.0 mM, to a protein solution (8.5 mg/mL), and the sample was concentrated to 39.0 mg/mL. 0.4 μL of the resulting protein−ligand solution was mixed with 0.2 μL of reservoir solution consisting of 19% (w/v) PEG 3350, 0.19 M sodium nitrate, and 0.1 M Bis-Tris pH 5.8. The plate was incubated at 20 °C, and crystals appeared within one month. The crystals were quickly transferred to a cryosolution consisting of 19% (w/v) PEG 3350, 0.19 M sodium nitrate, 0.1 M Bis-Tris pH 6.0, 21% (v/v) glycerol, 0.2 M sodium chloride, and 0.4 mM of **5** and then flash frozen in liquid nitrogen.

Crystallographic Data Collection, Phasing, and Refinement. Diffraction data were collected on the frozen crystals at the synchroton beamlines BL14.1 at BESSY (Berlin, Germany) and ID14.4 at ESRF (Grenoble, France). Data were indexed and integrated using the XDS package.[40] The ARTD8·**6** structure was solved by molecular replacement using Molrep[41] with the ARTD8 structure (PDB ID: 3SMI) as model template. The ARTD8·**5** structure was solved by molecular replacement using Molrep[41] with the ARTD8·**6** structure as model template. Similarly, the ARTD7·**1** structure was solved by molecular replacement with the ARTD7 *apo* structure (PDB ID: 3BLJ). Refinement was done using Refmac5[42] and Phenix.[43] Coot[44] was used for model building. The final models were validated and analyzed using Molprobity.[45] Data collection, phasing using molecular replacement, refinement procedures, and statistics are summarized in the Supporting Information.

Crystallographic Data Deposition. Coordinates and structure factors for the crystal structures have been deposited in the RCSB PDB,[46,47] under accession codes 4F0E, 4F1L, and 4F1Q.

Protein Structure Preparation for Virtual Screening. The D-loop in ARTD7 (ARTD7$^{459-656}$·**1**, monomer A, see Experimental Section details below) and ARTD8$^{1530-1720}$ (PDB ID: 3SMI, monomer A) was unaltered or truncated, resulting in two structures for each protein, namely, ARTD7$_{long}$, ARTD7$_{short}$, ARTD8$_{long}$, and ARTD8$_{short}$. ARTD7$_{long}$ included all residues in the D-loop. In ARTD7$_{short}$, 11 amino acids from Asn553 to Ser563 were removed. Five residues were missing from the D-loop in 3SMI: Gly1622, Lys1623, Asn1624, Ala1625, and Val1626. All remaining residues were kept in ARTD8$_{long}$. In ARTD8$_{short}$, six more residues were removed from the D-loop: Asn1617, Arg1618, Ser1619, Tyr1620, Ala1621, and Ala1627. Protein−structure preparations before docking (i.e. addition of hydrogens and sulfur bridges, optimization of hydroxyl, Asn, Gln, and His protonation states and tautomers, and removal of ligands) were performed in Maestro.[48]

Virtual Screening. A library of 17500 small molecules available from Laboratories for Chemical Biology Umeå, Chemical Biology Consortium Sweden, was filtered to provide a focused selection of molecules with at least one hydrogen-bond acceptor and one donor and a molecular weight of less than 600. This set was combined with seven known binders to give 8050 molecules that were docked to all four protein structures. A maximum of 16 tautomers and protonation states within the pH range 7 ± 2 were generated for each molecule using Epik.[49]

All dockings were performed using GlideSP.[50,51] Receptor grids were generated with identical dimensions for all four protein structures, with centerpoint coordinates x = −0.403, y = 10.413, and z = −28.813, and inner and outer box dimensions of 2744 and 3375 Å^3, respectively. The maximum number of poses per compound was 100, and the total number of poses was 1000000. No poses were rejected due to rms or placement similarities. A postdocking filter in Glide removed all poses that did not form a simultaneous hydrogen bond with the carbonyl oxygen and N-hydrogen in Gly538 in ARTD7 or Gly1602 in ARTD8.

All poses remaining from the docking and filtration were, in addition to glidescore, rescored with chemscore,[52] plp,[53] screenscore,[54] zapbind,[55] chemgauss3, and oechemscore, available in FRED,[56] and goldscore, available in GOLD.[57−59] The receptor grid in FRED and the docking region in GOLD used in the rescoring were comparable in size and placement to the docking grid used for docking in Glide. All score values were normalized with respect to the number of non-hydrogen atoms in the corresponding molecules, following the method presented by Pan et al.[60] to reduce the scoring bias of the scoring functions toward large molecules. The normalized score value (S_{norm}) was calculated according to

$$S_{norm} = S/N^{0.5} \quad (1)$$

where S is the original score value and N is the number of non-hydrogen atoms in the molecule.

Docking Score Resemblance and PCA. The overall binding capacity of the docked molecules for the ARTD proteins was determined by a docking-score evaluation approach based on normalized score values and PCA.[61] Five variables were calculated from the docking scores from the ensemble of poses generated for each molecule and for each of the eight scoring functions: the mean score, the standard deviation of the mean, the best score, the mean score of the 10 top-ranked poses, and the standard deviation of the top-ten mean. In addition, the number of poses (frequency) was recorded. All molecules with five poses or fewer (i.e., a frequency below six) were excluded from the PCA. The scoring data were mean centered and scaled to unit variance before PCA. PCA compresses data into fewer variables, principal components (PCs), which capture the main variation in the data.[62] The first PC (PC1) is oriented in the direction of the bulk of the variation in the data matrix. The values received by the molecules on these PCs are called *t*-scores, and the contributions from each of the original variables to the orientation of the PCs are called loadings.

A total of 41 variables, including the frequency, were subjected to PCA compression, resulting in four models, one for each virtual screen against one of the four protein structures. Strong outliers in the PCA models were detected and excluded iteratively based on their distance to the hyperplane spanned by the PCs (DModX) and/or their position in the PCA score space (see details in the Supporting Information). PCA modeling was performed using the SIMCA software.[63]

Selection of Molecules for Biological Evaluation Based on Docking Score Resemblance (Set A) and Glidescores (Set B). The selection of set A was based on the position of the known binders in the PCA PC space. All poses of the known binders that remained after the geometrical filtering (i.e., had the fork interaction with Gly) for each of the tautomers/protomers were included. The center coordinates for the known binders in the PC space were calculated, and these coordinates were used to calculate Euclidean distances from the center to each of the known binders. The radius of the volume within which the selection of molecules was performed was set to the average distance of the known binders to the center coordinate plus three standard deviations (corresponding to Euclidean distances of 4.98, 3.19, 4.77, and 5.83 for ARTD7$_{long}$, ARTD7$_{short}$, ARTD8$_{long}$, and ARTD8$_{short}$, respectively). Set A was selected to include all unique molecules from the screened compounds positioned within the established volume.

Set B was selected as follows. All poses from the virtual screen against the four proteins were sorted according to normalized (eq 1) glidescore top-score values, and a number of the highest ranked unique molecules (i.e., tautomer- and multiplet-independent), matching the number selected for set A, were picked for biological evaluation.

Descriptor Calculations and Clustering. Physicochemical descriptors and MACCS fingerprints[35] were calculated for all virtual screening hits using the MOE software.[64] The hits for ARTD7 and ARTD8 were clustered based on their fingerprint profile similarity using Tanimoto coefficients with a cutoff of 60 for similarity within and overlap between clusters.[64]

Compounds. All compounds used for crystallization, differential scanning fluorimetry, and isothermal titration calorimetry were either obtained from Laboratories for Chemical Biology Umeå, Chemical Biology Consortium Sweden, or synthesized as described below. The purity of commercial compounds were confirmed to be >95% by a reversed-phase HPLC UV-trace using a C_{18}, 5 μm, column and a H_2O/acetonitrile eluent system. The structures of commercial compounds were confirmed using ^{1}H NMR spectroscopy [400 MHz in $(CD_3)_2SO$], and spectra are presented in the Supporting Information. High-resolution mass spectrometry (HRMS) was performed with electrospray ionization (ES+) using sodiumformate as calibration chemical.

Differential Scanning Fluorimetry. Proteins were diluted to 0.2 mg/mL in buffer [10 mM sodium acetate pH 5.5 and 2 mM tris(2-carboxyethyl)phosphine (TCEP) for ARTD8, PBS and 2 mM TCEP for ARTD7, and 20 mM HEPES pH 7.5, 300 mM NaCl, and 2 mM TCEP for ARTD1] and 1:1000 dilution of SyproOrange (Invitrogen). Protein solution (24.5 μL) was added to all wells in clear 96-well plates (BioRad), each well containing 0.5 μL of predispensed compound solution (2.5 mM in DMSO). Optical tape (Bio-Rad) was used to seal the plates. Thermal stability was measured by monitoring SyproOrange fluorescence ($\lambda_{excitation}$ = 490 nm and $\lambda_{emission}$ = 575 nm) while heating the samples from 20 to 90 °C in increments of 1 °C/min in a Bio-Rad iCycler. The thermal shifts (ΔT_m) were determined as described previously,[65] using pure DMSO (2% v/v) as reference. Compounds were measured in triplicate, with the exception of the compounds selected from glidescore scoring, which were measured in duplicate. Three compounds were excluded owing to their poor solubility.

Synthesis of Compounds 5 and 6. *(E)-4-(3-Carbamoylphenylamino)-4-oxobut-2-enoic Acid (**5**).* Fumaric acid (**27**) (426 mg, 3.67 mmol) was suspended in dry CH_2Cl_2 (4 mL) and oxalyl chloride (0.64 mL, 7.34 mmol) was added dropwise over 5 min. Dimethylformamide (1 drop) was added, and the mixture was stirred at 40 °C for 7 h. Solvent and excess oxalyl chloride was removed in vacuo, and the residue was redissolved in dioxane (6 mL). 3-Aminobenzamide (100 mg, 0.73 mmol) in dioxane (4 mL) was added dropwise over 5 min at

dx.doi.org/10.1021/jm300746d | J. Med. Chem. 2012, 55, 7706−7718

rt, and the reaction was stirred at 40 °C for 16 h. The reaction mixture was diluted with CH_2Cl_2 and extracted with NaOH (1 M) and adjustment of the water phase pH to ∼1 using HCl (1 M). The precipitate was collected by filtration and purified by HPLC (H_2O:MeCN 90:10 to 30:70 over 40 min) to give (*E*)-4-(3-carbamoylphenylamino)-4-oxobut-2-enoic acid (71 mg, yield 41%). ^{1}H NMR (400 MHz, $(CD_3)_2SO$) δ 13.00 (s, 1H), 10.63 (s, 1H), 8.12 (s, 1H), 7.95 (s, 1H), 7.84 (d, J = 8.1 Hz, 1H), 7.59 (d, J = 8.1 Hz, 1H), 7.41 (t, J = 8.0 Hz, 1H), 7.36 (s, 1H), 7.15 (d, J = 15.5 Hz, 1H), 6.68 (d, J = 15.5 Hz, 1H). ^{13}C NMR (100 MHz, $(CD_3)_2SO$) δ 167.7, 166.2, 161.7, 138.6, 137.0, 135.2, 130.9, 128.7, 122.8, 122.1, 119.0. See the Supporting Information for ^{1}H NMR and ^{13}C NMR spectra. HRMS (ES^+) calcd [M + Na^+] for $C_{11}H_{10}N_2NaO_4^+$ 257.0538, obsd 257.0538.

*(Z)-4-(3-Carbamoylphenylamino)-4-oxobut-2-enoic Acid (**6**).* 3-Aminobenzamide (**28**) (50 mg, 0.37 mmol) and maleic anhydride (43 mg, 0.44 mmol) were dissolved in tetrahydrofuran (1 mL) and stirred for 18 h at rt. The solid material was collected by filtration and dried in vacuo to give (*Z*)-4-(3-carbamoylphenylamino)-4-oxobut-2-enoic acid (71 mg, yield 83%). ^{1}H NMR [400 MHz, $(CD_3)_2SO$] δ 13.04 (s, 1H), 10.50 (s, 1H), 8.08 (t, J = 1.9 Hz, 1H), 7.94 (s, 1H), 7.78 (dd, J = 8.0, 2.3 Hz, 1H), 7.57 (dt, J = 7.7, 1.7 Hz, 1H), 7.39 (t, J = 7.9 Hz, 1H), 7.34 (s, 1H), 6.47 (d, J = 12.1 Hz, 1H), 6.32 (d, J = 12.1 Hz, 1H). ^{13}C NMR [100 MHz, $(CD_3)_2SO$] δ 167.7, 166.9, 163.3, 138.6, 135.2, 131.5, 130.5, 128.6, 122.5, 122.1, 119.1. HRMS (ES^+) calcd [M + Na^+] for $C_{11}H_{10}N_2NaO_4^+$ 257.0538, obsd 257.0540.

Isothermal Titration Calorimetry. The ITC experiments were performed using ARTD8 expressed and purified according to the protocol described in this paper. All buffers were degassed and sterile filtered prior to use. Solutions of 0.750 mM of **6** and 1.125 mM of **5** were prepared in PBS buffer pH 7.4, (50 mM phosphate buffer, 150 mM sodium chloride) with 2 mM TCEP. The ITC experiments were performed at 25.0 °C (298.15 K) using an ITC200 (MicroCal, GE Healthcare Life Sciences). Protein solutions were diluted in the PBS buffer mentioned above (56.1 μM for **6** and 87.2 μM for **5**) and were loaded in the calorimeters' sample cell. The compound solutions were loaded in the syringe, and 35 automated injections of 1.11 μL each with a 250 s break in between each injection were made. The stirring speed was 600 rpm. The injection heat was automatically integrated, and the data were analyzed using the single-site binding model, with 300 iterations of the Chi-Sqr fitting routine in Origin SR4 software.[66]

Molecular Dynamics Simulations. The structures of the ARTD8·**6** and ARTD8·**5** complexes used in the MD simulations originated from crystal structures presented in this work (PDB IDs: 4F1L and 4F1Q). In the cases of ARTD8 structures with incomplete D-loops (i.e., monomer B and D in 4F1L and monomer A and B in 4F1Q), the loops were modeled in open and closed forms using the program Coot.[44] The modeling of the protein backbone was guided by ARTD8 structures with intact D-loops (PDB IDs 3SMJ and 4F1L). Missing side-chains were added, and the rotamers with best fit were selected. MD simulations were carried out using AMBER[67,68] with the AMBER 1999SB force field[69] and the general AMBER force field GAFF[70] for the proteins and ligands, respectively. GAFF parameters for ligands with AM1 partial charges were prepared in the AMBER "parmchk" module. The protein−ligand complex was solvated by adding TIP3P waters to constitute an octahedral water box with dimensions 88 Å × 88 Å × 88 Å. The distance between the outer boundary of the water box, and the solute surface was set to 10 Å. Relaxation, temperature equilibration, and MD routines were conducted using the AMBER Sander module (see Supporting Information for further details). During the temperature equilibration and MD routines, a nonbonded cutoff distance of 9 Å was applied to handle electrostatic interactions in periodic boxes by the Particle Mesh Ewald method.[71] The SHAKE method[69] was applied to ensure constant proton−heteroatom bond lengths. Coordinates (frames) and energy outputs from the relaxation and the MD routines were saved every 200 iterations (every 0.4 ps). Finally, 5 ns simulations were conducted for the single-monomer ARTD8·**5** and ARTD8·**6** complexes.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Multiple sequence alignment of the D-loop for the ARTD family of proteins; structures and binding data for known binders to ARTD7 and ARTD8; number of molecules post filtration and PCA details; $ARTD8_{long}$ PCA *t*-score and loading plots for PCs 3−5; ^{1}H NMR spectra for virtual screening hits; examples of docking poses; compounds that bound to ARTD1 (and not to ARTD7−8) with a ΔT_m-shift ≥1.5 °C; the crystal packing and unit cell for ARTD8 structures; cloning and protein expression; protein purification; crystallographic data collection, phasing, and refinement statistics; ^{1}H and ^{13}C NMR spectra for synthesized compounds; description of MD simulations; rmsd and distance plots from single monomer MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

PDB accession codes: 4F0E, 4F1L, 4F1Q.

■ AUTHOR INFORMATION

Corresponding Author

*for A.L.: phone, +46907866890; E-mail, anna.linusson@chem.umu.se. For H.S.: E-mail: herwig.schuler@ki.se.

Present Address

§Royal Swedish Academy of Engineering Sciences, Box 5073, SE-102 42 Stockholm, Sweden.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Swedish Foundation for Strategic Research (Stiftelsen för strategisk forskning, SSF) for provided financial support, Laboratories for Chemical Biology Umeå, Chemical Biology Consortium Sweden, for access to the small-molecule collection, the beamline staff at BESSY and ESRF synchrotron radiation facilities for excellent support, the Protein Science Facility at Karolinska Institutet (psf.ki.se) for excellent support, and the High Performance Computing Center North (HPC2N) (www.hpc2n.umu.se) for computer resources. A.L., M.E., and P.W.S. thank the Swedish Research Council for provided financial support. P.W.S. and A.L. thank Umeå University for provided financial support. P.W.S. thanks the Knut and Alice Wallenberg Foundation and Göran Gustafsson Foundation for provided financial support.

■ ABBREVIATIONS USED

ARTD, diphtheria toxin-like ADP-ribosyltransferases; PARP, poly(ADP-ribose) polymerase; NAD^+, nicotinamide adenine dinucleotide; PDB, Protein Data Bank; ITC, isothermal titration calorimetry; PCR, polymerase chain reaction; PCA, principal component analysis; PC, principal component; rmsd, root-mean-square distance; MD, molecular dynamics; HRMS, high-resolution mass spectrometry

■ REFERENCES

(1) Hottiger, M. O.; Hassa, P. O.; Lüscher, B.; Schüler, H.; Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **2010**, *35*, 208−219.

(2) Hassa, P. O.; Hottiger, M. O. The diverse biological roles of mammalian PARPS, a small but powerful family of poly-ADP-ribose polymerases. *Front. Biosci.* **2008**, *13*, 3046−3082.

(3) Underhill, C.; Toulmonde, M.; Bonnefoi, H. A review of PARP inhibitors: from bench to bedside. *Ann. Oncol.* **2011**, *22*, 268−279.
(4) Aguiar, R. C. T.; Takeyama, K.; He, C. Y.; Kreinbrink, K.; Shipp, M. A. B-Aggressive lymphoma family proteins have unique domains that modulate transcription and exhibit poly(ADP-ribose) polymerase activity. *J. Biol. Chem.* **2005**, *280*, 33756−33765.
(5) Kleine, H.; Poreba, E.; Lesniewicz, K.; Hassa, P. O.; Hottiger, M. O.; Litchfield, D. W.; Shilton, B. H.; Lüscher, B. Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. *Mol. Cell* **2008**, *32*, 57−69.
(6) Goenka, S.; Cho, S. H.; Boothby, M. Collaborator of stat6 (CoaSt6)-associated poly(ADP-ribose) polymerase activity modulates stat6-dependent gene transcription. *J. Biol. Chem.* **2007**, *282*, 18732−18739.
(7) Cho, S. H.; Ahn, A. K.; Bhargava, P.; Lee, C. H.; Eischen, C. M.; McGuinness, O.; Boothby, M. Glycolytic rate and lymphomagenesis depend on PARP14, an ADP ribosyltransferase of the B aggressive lymphoma (BAL) family. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15972−15977.
(8) Frye, S. V. The art of the chemical probe. *Nature Chem. Biol.* **2010**, *6*, 159−161.
(9) Wahlberg, E.; Karlberg, T.; Kouznetsova, E.; Markova, N.; Macchiarulo, A.; Thorsell, A.; Pol, E.; Frostell, Å.; Ekblad, T.; Öncü, D.; Kull, B.; Robertson, G.; Pellicciari, R.; Schüler, H.; Weigelt, J. Family wide chemical profiling and structural analysis of PARP and Tankyrase inhibitors. *Nature Biotechnol.* **2012**, *30*, 283−288.
(10) Bell, C. E.; Eisenberg, D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **1996**, *35*, 1137−1149.
(11) Ruf, A.; de Murcia, G.; Schulz, G. E. Inhibitor and NAD(+) binding to poly(ADP-ribose) polymerase as derived from crystal structures and homology modeling. *Biochemistry* **1998**, *37*, 3893−3900.
(12) Lehtiö, L.; Jemth, A. S.; Collins, R.; Loseva, O.; Johansson, A.; Markova, N.; Hammarström, M.; Flores, A.; Holmberg-Schiavone, L.; Weigelt, J.; Helleday, T.; Schüler, H.; Karlberg, T. Structural basis for inhibitor specificity in human poly(ADP-ribose) polymerase-3. *J. Med. Chem.* **2009**, *52*, 3108−3111.
(13) Karlberg, T.; Hammarström, M.; Schütz, P.; Svensson, L.; Schüler, H. Crystal structure of the catalytic domain of human PARP2 in complex with PARP inhibitor ABT-888. *Biochemistry* **2010**, *49*, 1056−1058.
(14) Karlberg, T.; Markova, N.; Johansson, I.; Hammarström, M.; Schütz, P.; Weigelt, J.; Schüler, H. Structural basis for the interaction between tankyrase-2 and a potent wnt-signaling inhibitor. *J. Med. Chem.* **2010**, *53*, 5352−5355.
(15) Narwal, M.; Venkannagari, H.; Lehtiö, L. Structural basis of selective inhibition of human tankyrases. *J. Med. Chem.* **2012**, *55*, 1360−1367.
(16) Sun, J. J.; Maresso, A. W.; Kim, J. J. P.; Barbieri, J. T. How bacterial ADP-ribosylating toxins recognize substrates. *Nature Struct. Mol. Biol.* **2004**, *11*, 868−876.
(17) Branca, D.; Cerretani, M.; Jones, P.; Koch, U.; Orvieto, F.; Palumbi, M. C.; Rowley, M.; Toniatti, C.; Muraglia, E. Identification of aminoethyl pyrrolo dihydroisoquinolinones as novel poly(ADP-ribose) polymerase-1 inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4042−4045.
(18) Zakharenko, A. L.; Sukhanova, M. V.; Khodyreva, S. N.; Novikov, F. N.; Stroylov, V. S.; Nilov, D. K.; Chilov, G. G.; Svedas, V. K.; Lavrik, O. I. Improved procedure of the search for poly(ADP-ribose) polymerase-1 potential inhibitors with the use of the molecular docking approach. *Mol. Biol.* **2011**, *45*, 517−521.
(19) Novikov, F. N.; Stroylov, V. S.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. Developing novel approaches to improve binding energy estimation and virtual screening: a PARP case study. *J. Mol. Model.* **2009**, *15*, 1337−1347.
(20) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein−ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851−5855.
(21) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580−594.
(22) Kirchmair, J.; Distinto, S.; Schuster, D.; Spitzer, G.; Langer, T.; Wolber, G. Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates. *Curr. Med. Chem.* **2008**, *15*, 2040−2053.
(23) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein−ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.
(24) Feher, M. Consensus scoring for protein−ligand interactions. *Drug Discovery Today* **2006**, *11*, 421−428.
(25) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein−ligand complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380−391.
(26) Lehtiö, L.; Collins, R.; van den Berg, S.; Johansson, A.; Dahlgren, L. G.; Hammarström, M.; Helleday, T.; Holmberg-Schiavone, L.; Karlberg, T.; Weigelt, J. Zinc binding catalytic domain of human tankyrase 1. *J. Mol. Biol.* **2008**, *379*, 136−145.
(27) Wagner, G.; Krasselt, U. Multicyclic azines with heteroatoms in position-1 and 0.38. Synthesis of Pyrido[3′,2′,/4,5]thienol[3,2-D]-pyrimidines with a mercapto-substituted 1,2,4-triazole ring anellated differently on the pyrimidine ring. *Pharmazie* **1992**, *47*, 943−943.
(28) Eldin, N. K. Reactions of phthalimidobenzoic acid azides with trimethyl phosphite and tris (dialkylamino) phosphines. *Phosphorus, Sulfur Silicon Relat. Elem.* **1995**, *102*, 25−29.
(29) Munchhof, M. J.; Sobolov-Jaynes, S. B. Thienopyrimidine and thienopyridine derivatives useful as anticancer agents. Patent WO9924440 (A1), 1999.
(30) Al-Mawsawi, L. Q.; Sechi, M.; Neamati, N. Single amino acid substitution in HIV-1 integrase catalytic core causes a dramatic shift in inhibitor selectivity. *FEBS Lett.* **2007**, *581*, 1151−1156.
(31) Wagner, G.; Briel, D.; Leistner, S. Ring rearrangement reaction of a 1,2,4-dithiazolidine with hydroxylamine or hydrazine yielding a 1,2,4-oxadiazole or a 1,2,4-triazole derivative. *Pharmazie* **1980**, *35*, 48−49.
(32) Burckhalter, J. H.; Dill, D. R. 8-(Dialkylaminoalkylamino)-caffeines as potential pharmacodynamic agents. *J. Am. Pharm. Assoc.* **1959**, *48*, 190−194.
(33) Oprea, T. I.; Allu, T. K.; Fara, D. C.; Rad, R. F.; Ostopovici, L.; Bologa, C. G. Lead-like, drug-like or "pub-like": how different are they? *J. Comput.-Aided Mol. Des.* **2007**, *21*, 113−119.
(34) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743−3748.
(35) *MACCS Keys*; MDL Information Systems, Inc.: 14600 Catalina Street, San Leandro, CA 94577.
(36) Zavodszky, M. I.; Rohatgi, A.; Van Voorst, J. R.; Yan, H. G.; Kuhn, L. A. Scoring ligand similarity in structure-based virtual screening. *J. Mol. Recognit.* **2009**, *22*, 280−292.
(37) Andersson, C. D.; Chen, B. Y.; Linusson, A. Multivariate assessment of virtual screening experiments. *J. Chemom.* **2010**, *24*, 757−767.
(38) Compound **6** bound to ARTD1 with a ΔT_m of 0.36 and compound **5** bound to ARTD1 with a ΔT_m of 0.20.
(39) Compounds **5** and **6** had a water buffer solubility of at least 2.5 mM, determined by ocular inspection during ITC experiments.
(40) Kabsch, W. Xds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 125−132.
(41) Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, *30*, 1022−1025.
(42) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *D53*, 240−255.
(43) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H.

PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 213−221.
(44) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 486−501.
(45) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *D66*, 12−21.
(46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(47) Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB), www.pdb.org.
(48) *Maestro, 9.1*; Schrödinger, LLC; 120 West 45th Street, 17th Floor, Tower 45, New York, NY; pp 10036−4041.
(49) *Epik, 2.1207*; Schrödinger, LLC: 120 West 45th Street, 17th Floor, Tower 45, New York, NY; pp 10036−4041.
(50) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(51) *Glide, 5.6109*; Schrödinger, LLC: 120 West 45th Street, 17th Floor, Tower 45, New York, NY; pp 10036−4041.
(52) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions 0.1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(53) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Deciphering common failures in molecular docking of ligand−protein complexes. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 731−751.
(54) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(55) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson−Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608−640.
(56) *FRED, 2.2.5*; Openeye Scientific Software Inc.: 3600 Cerrillos Road, Suite 1107, Santa Fe, NM 87507.
(57) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.
(58) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(59) *GOLD, 4.0.1*; The Cambridge Crystallographic Datacenter: 12 Union Road, Cambridge, CB2 1EZ, U.K.
(60) Pan, Y. P.; Huang, N.; Cho, S.; MacKerell, A. D. Consideration of molecular weight during compound selection in virtual target-based database screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267−272.
(61) Andersson, I. E.; Andersson, C. D.; Batsalova, T.; Dzhambazov, B.; Holmdahl, R.; Kihlberg, J.; Linusson, A. Design of glycopeptides used to investigate class II MHC binding and T-cell responses associated with autoimmune arthritis. *PLoS One* **2011**, *6*.
(62) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(63) *SIMCA-P+, 12.0*; Umetrics AB: Box 7960, Umeå, Sweden.
(64) *MOE (The Molecular Operating Environment)*; Chemical Computing Group Inc.: 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7, 2010; p 10.
(65) Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protoc.* **2007**, *2*, 2212−2221.
(66) *Origin SR4, 7.0552*; Origin Lab Corperation: One Roundhouse Plaza, Northampton, MA 01060, USA.
(67) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(68) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, 2010.
(69) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(70) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(71) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.