ARTICLE

ACS chemical biology

Targeting Multifunctional Proteins by Virtual Screening: Structurally Diverse Cytohesin Inhibitors with Differentiated Biological Functions

Dagmar Stumpfe[†], Anke Bill[‡], Nina Novak[§], Gerrit Loch[∥], Heike Blockus[‡], Hanna Geppert[†], Thomas Becker[∥], Anton Schmitz[‡], Michael Hoch[∥], Waldemar Kolanus[§], Michael Famulok[‡,*], and Jürgen Bajorath[†,*]

[†]Chemical Biology & Medicinal Chemistry, c/o Department of Life Science Informatics, B-IT, Dahlmannstr. 2, 53113 Bonn, Germany, [‡]Chemical Biology & Medicinal Chemistry, c/o Kekulé Institute of Organic Chemistry and Biochemistry, Gerhard-Domagk-Strasse 1, 53121 Bonn, Germany, [§]Laboratory of Molecular Immunology, Carl-Troll-Str. 31, 53115 Bonn, Germany, and [∥]Development and Genetics, Laboratory of Molecular Developmental Biology, Carl-Troll-Str. 31, University of Bonn, LIMES Institute, 53115 Bonn, Germany

ABSTRACT Virtual screening (VS) of chemical libraries formatted *in silico* provides an alternative to experimental high-throughput screening (HTS) for the identification of small molecule modulators of protein function. We have tailored a VS approach combining fingerprint similarity searching and support vector machine modeling toward the identification of small molecular probes for the study of cytohesins, a family of cytoplasmic regulator proteins with multiple cellular functions. A total of 40 new structurally diverse inhibitors were identified, and 26 of these compounds were more active than the primary VS template, a single known inhibitory chemotype, in at least one of three different assays (guanine nucleotide exchange, *Drosophila* insulin signaling, and human leukocyte cell adhesion). Moreover, these inhibitors displayed differential inhibitory profiles. Our findings demonstrate that, at least for the cytohesins, computational extrapolation from known active compounds was capable of identifying small molecular probes with highly diversified functional profiles.

High-throughput screening (HTS) currently plays a major role as a source of novel active molecules that serve as leads for drug development and as tools for biomedical research. Furthermore, a variety of virtual screening (VS) methods including target structure- (*1*) and ligand-based approaches (*2*) can also be employed to aid in the identification of active compounds. However, all of these methods are limited in their accuracy, and to this date it has been essentially impossible to rationalize why certain methods succeed or fail on a given target (*3*). Hence, successful VS applications are currently far from being routine. Nevertheless, VS methods are widely applied, and a plethora of approaches of often very different computational complexity are being investigated for hit identification (*3*).

Ligand-based VS methods utilize information from known active compounds as input, with the primary goal to abstract from these reference compounds and identify structurally diverse hits (*2, 3*). Most ligand-based methods require multiple active (and often also inactive) compounds as references, and it is generally assumed that the probability of VS success substantially increases with the amount of available compound and structure−activity relationship (SAR) data (*2, 3*). However, in a few instances, the use of single reference compounds has also resulted in the identification of novel hits (*4*). Failures in VS are often attributed to the availability of only limited SAR information, in addition to lim-

***Corresponding authors,**
bajorath@bit.uni-bonn.de,
m.famulok@uni-bonn.de.

Received for review June 29, 2009 and accepted July 8, 2010.

Published online July 8, 2010

10.1021/cb100171c

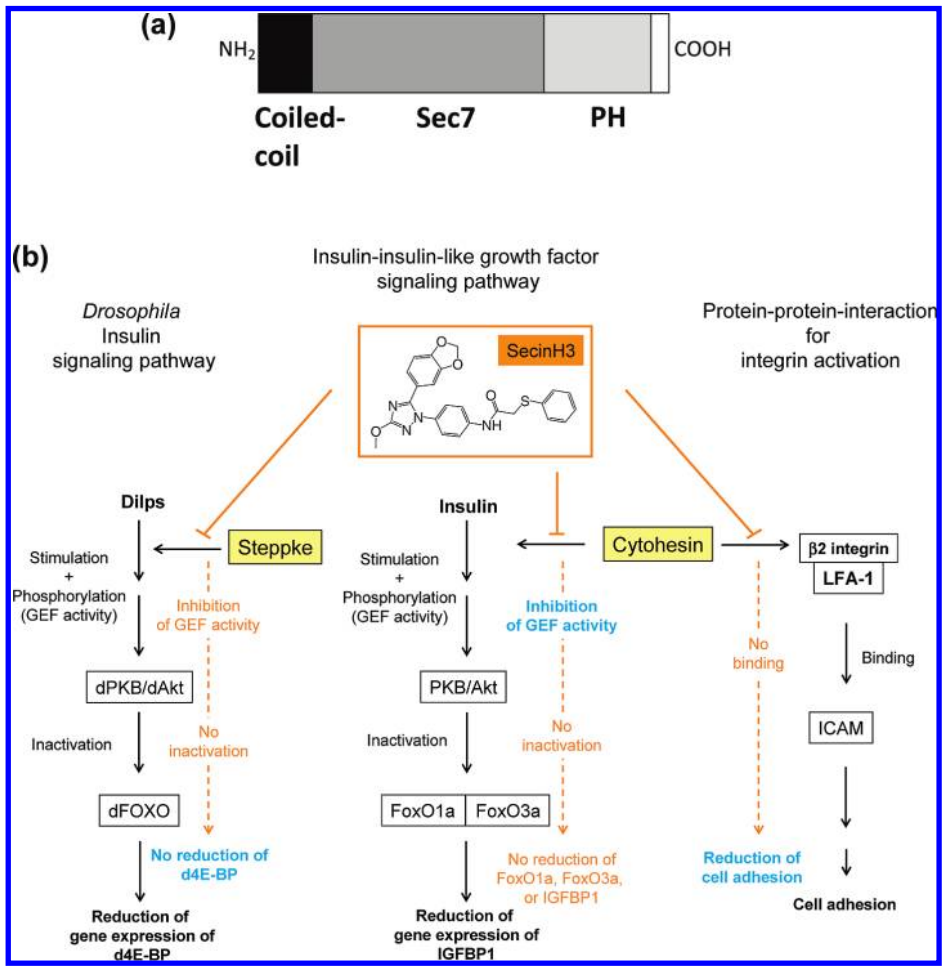

Figure 1. Cytohesin domain structure and functions. a) The domain organization of the cytohesins. The N-terminal coiled-coil region is followed by the Sec-7 and a PH domain. The GEF and cell adhesion functions reside in the Sec-7 domain (see text). b) Involvement of cytohesin in the insulin-signaling pathway in human cells and in *Drosophila*, together with its role in integrin activation and cell adhesion, and the structure of the *pan*-cytohesin inhibitor SecinH3. Cytohesin-mediated cell adhesion does not involve guanine nucleotide exchange activtiy. Black arrows represent native signaling pathways, and orange arrows represent pathways impaired through cytohesin/Steppke inhibition. Biological assay used in our study monitored pathway stages highlighted in blue.

ited accuracy of VS algorithms. Even in successful cases, there are other complications. For example, structurally diverse compounds identified through ligand-based VS are generally less potent than reference molecules because they are selected to structurally depart from known reference molecules that are often highly optimized structural motifs (*3*).

Search calculations typically focus on compounds that mimic a specific biological activity, and this "single target−single activity" paradigm represents a conceptual cornerstone of VS efforts. Hence, in addition to open methodological questions, another largely unexplored topic in computational screening is targeting of multifunctional proteins. There is currently only very little, if any, information available as to whether it is possible through VS to identify small molecular probes that modulate multiple cellular functionalities of target proteins.

Here, we have set out to investigate this issue by targeting a class of cytoplasmic proteins, the cytohesins, which present a *par excellence* example of a multifunctional protein family. Cytohesins are small guanine nucleotide exchange factors (GEFs) that stimulate ADP ribosylation factors, ubiquitously expressed Ras-like GTPases, which control various cellular regulatory networks ranging from vesicle trafficking and integrin activation to insulin signaling (*5*–*8*). Members of the cytohesin family include cytohesin-1; cytohesin-2 (ARNO); cytohesin-3, also known as Grp-1 in humans and Steppke in *Drosophila* (*6*); and cytohesin-4 (*7*). All four currently known cytohesins share a conserved multidomain structure, *i.e.*, a Sec-7 domain harboring the GEF activity, a pleckstrin-homology- (PH), and a coiled-coil domain, as illustrated in Figure 1, panel a. Cytohesin-1 and -2 were originally identified on the basis of the ability of their Sec-7 domain to interact with the cytoplasmic β-2 chain of the integrin leukocyte function-associated antigen-1 (LFA-1) β-2 (*8*), which implicated cytohesin in β-2 integrin-mediated inside-out signaling and immune cell adhesion. In inside-out signaling, a number of cytoplasmic proteins, including cytohesin-1 and -2, aid in altering the conformations of the extracellular domains of integrins, thus enabling their binding to cognate ligands (*8, 9*). Leukocytes in particular depend on this process in order to leave the vasculature and enter lymph nodes or inflamed tissues.

Cytohesin-1 was subsequently shown to be involved in mitogen-associated protein kinase signaling in tumor cell proliferation as well as in T-helper cell activation and differentiation (*10, 11*). Furthermore, cytohesin-3 was very recently identified as an essential component of the phosphatidylinositol-3-kinase (PI3K) pathway in insulin signal transduction in human liver cells, *Drosophila*, and mouse (*6, 12*). By contrast, cytohesin-4 has not yet been functionally characterized.

Although small molecular probes that effectively interfere with different cytohesin functions are highly desirable, only one inhibitory chemotype has thus far been described, the *pan*-active cytohesin inhibitor SecinH3, a 1-, 2-, 4-substituted triazole (Figure 1, panel b), which has been shown to target the Sec-7 domain of cytohesins-1, -2, and -3 and inhibit their GEF activity and cytohesin-associated functions including insulin signaling (Figure 1, panel b) (*12*). Insulin- or, in the case of *Drosophila melanogaster*, *Drosophila* insulin-like peptides (Dilps)-stimulation leads to activation of protein kinase B (vertebrates, PKB/Akt; *Drosophila*, pPKB/dAkt) through phosphorylation. PKB/pPKB then translocates into the nucleus where it phosphorylates the forkhead-box transcription factors FoxO1a and FoxO3a in vertebrates or dFOXO in *Drosophila*, leading to their deactivation and nuclear exclusion. FoxO1a and FoxO3a regulate the expression of the insulin-like growth factor binding protein 1 (IGFBP1), whereas dFOXO regulates the gene expression of d4E-BP, Steppke, and the insulin receptor (*6*). SecinH3 inhibits the cytohesin/Steppke-dependent activation of PKB/Akt and pPKB/dAkt, respectively, which results in an increase of nuclear FoxO1a, FoxO3a, and dFOXO levels and in transcription of IGFBP1 (*12, 13*) and d4E-BP (*6*), respectively.

Cytohesin-1 and -2 also interact with the cytoplasmic sequence of the LFA-1 β2 chain, CD18, which reflects a role of these proteins in signal complex assembly with the cytoplasmic tail of β2 integrins (*8*). This interaction results in an increased avidity of integrin−substrate binding, and thereby cytohesins mediate β2 integrin-dependent adhesion of T-cells. The GEF activity of cytohesin-1 appears to play a minor role, if any, in this activity, and T-cell adhesion was only weakly inhibited by SecinH3.

In addition to our motivation to explore the potential of VS approaches on multifunctional proteins using cytohesins as a test case, the computational search for cytohesin inhibitors has also high practical relevance. Owing to the complexity of cytohesin functional assays and the ensuing limited throughput capacity, an HTS campaign covering the various biological activities of cytohesins is difficult to conduct. We therefore have designed a VS protocol to extrapolate from the SecinH3 triazole chemotype, to explore its extended chemical neighborhood, and to identify structurally diverse molecules that gradually depart from it. However, a principal caveat (as discussed above) has been that in this case only very limited ligand information was available as a starting point for VS. Thus, the investigation was considered a challenging test case.

RESULTS AND DISCUSSION

Virtual Screening Strategy. For our cytohesin inhibitor search, we have designed a VS strategy utilizing methods that had to meet three criteria, *i.e.*, the methods should be (i) capable of operating on the basis of only limited reference information (SecinH3 and a few related compounds), (ii) based on chemically intuitive and simple molecular representations, and (iii) widely available. Therefore, two standard methodologies were combined. Chemical similarity searching using three 2D molecular fingerprints (FPs) of distinct design was carried out to rank database compounds in the order of decreasing molecular similarity to SecinH3 and to select candidate compounds ranging from SecinH3 analogues to distantly related structures. Accordingly, in the following, this approach is referred to as a "similarity search". Furthermore, in order to focus the search on increasingly diverse molecules, we also conducted support vector machine (SVM) calculations using 2D molecular fingerprints as descriptors. Thus, for both fingerprint and SVM searching, reference information was exclusively extracted from simple 2D molecular graphs of reference molecules. For SVM analysis, SecinH3 and only seven other compounds including SecinH3 analogues and in-house hits from the original aptamer screening assay were used as positive training set compounds, and 1000 randomly selected database compounds were used as negative training examples. For machine learning methods, this training set is rather limited in size. Two SVM models based on different molecular representations (descriptors) were developed and applied to predict active database compounds and rank them. From SVM calculations, only compounds with chemical scaffolds distinct from that of SecinH3 were selected.

Accordingly, this approach is referred to as a "diversity search".

From each individual FP search list, the top 20 candidate molecules were selected, and from each SVM ranking, the top 50 candidates. In addition, the top 1000 molecules were compared in FP and SVM rankings, and any compound occurring in at least two FP or two SVM lists was also selected. This protocol resulted in the preselection of 169 similarity search and 114 diversity search candidates, and 145 of these compounds could be acquired from commercial sources.

Biological Evaluation. We have used three different biological assays to test our 145 candidate compounds for their ability to interfere with cytohesin functions. Isolated ARNO-Sec-7 and ARF1 were used in a guanine nucleotide exchange assay in order to identify candidate compounds inhibiting cytohesin GEF activity. In addition, in *Drosophila* S2 tissue culture cells, compounds were tested for their ability to interfere with cytohesin-dependent insulin signaling, which results in the transcriptional activation of dFOXO-dependent target genes such as 4EB-P (Figure 1). Furthermore, the potential of candidate compounds was evaluated to block cytohesin-mediated cell adhesion of human leukocytes.

The initial screen of our 145 candidates revealed that ~40 compounds showed measurable activity in one or more of our assays. We then focused on those inhibitors that were more active than SecinH3. Accordingly, for classifying a candidate compound as a hit, the following criteria were applied: at least 10% stronger inhibition than SecinH3 in guanine nucleotide exchange assays, at least 50% reduction relative to SecinH3 in cell adhesion assays using either PMA- or OKT3-stimulated cells, or at least a 2-fold increase in dFOXO-dependent 4E-BP levels. On the basis of these criteria, 26 active compounds were selected and confirmed as hits. These compounds, their primary activity, and source information are reported in Table 1, and assay results are shown in Figure 2. Because all of these inhibitors were active in nucleotide exchange assays (Figure 2, panel a), these compounds bound to the Sec-7 domain.

Fifteen compounds were more active than SecinH3 in nucleotide exchange assays, seven in dFOXO, and 11 in adhesion assays. For individual compounds, improvements in inhibition levels over SecinH3 were in part substantial. For example, compound 16 (consecutive number on the 145 candidate selection list; no rank information), termed Secin16, induced a ~20-fold increase in dFOXO-dependent 4E-BP transcript levels compared to SecinH3 and was the by far most active compound in *Drosophila* assays. The IC_{50} value of Secin16 measured for the isolated Sec-7 domain of cytohesin-2 was 3.1 μM compared to 11.4 μM for SecinH3. With an IC_{50} value of 2.1 μM, Secin69 was slightly more active than Secin16, and the most active new inhibitor in nucleotide exchange assays. Both Secin16 and 69 inhibition displayed clear dose−response behavior (Figure 2, panel b), indicating specificity of the interactions. Secin69 was also active in cell adhesion assays. However, the apparent inhibitory effect of Secin69 in cell adhesion was due to strong toxicity of the compound on Jurkat cells, as assessed by Trypan Blue staining, in contrast to Secin16. In cell adhesion assays, Secin107 was most active, reducing adhesion to only ~9% of the SecinH3 level. Secin16 and 132 were the only two compounds with a better inhibition compared to that of SecinH3 in all three assays.

Because all inhibitors were active in nucleotide exchange assays and dose−response behavior was observed, binding to the Sec-7 domain was specific. For the overall most promising compound, Secin16, we also confirmed Sec-7 domain binding by surface plasmon resonance (using a novel chip design and detection method; see Methods). Experiments were performed measuring binding and dissociation of Secin16 on the immobilized cytohesin-1 Sec-7 domain at different concentrations. As shown in Figure 2, panel e, binding was concentration-dependent and saturation of binding was observed, resulting in an estimated K_D of 7.5 μM. Under the same experimental conditions, saturation of binding was difficult to achieve for SecinH3, consistent with weaker binding. We further confirmed the interaction of Secin16 with the Sec7 domain by performing microscale thermophoresis measurements (*27, 28*). In good agreement with the SPR data, a K_D of 5 ± 1 μM for the binding of Secin16 to ARNO-Sec7 was obtained. The K_D for SecinH3 was higher than for Secin16 but could not be quantified due to the limited solubility of SecinH3 under these conditions, which prevented saturation of binding.

Figure 3 compares SecinH3, Secin16, 69, 107, and 132. Secin69, 16, and 132 are more potent inhibitors in nucleotide exchange assays than SecinH3 (IC_{50} 11.4 μM), with IC_{50} values of 2.1, 3.1, and 8.0 μM, respectively, whereas Secin107 is slightly less potent (14.9 μM).

TABLE 1. Confirmed active compounds[a]

ZINC ID	Compound	Hit	Strategy[b]	Methodology[c]	Supplier and code
ZINC00843734	Secin16	Consensus	Diversity	SVM GpiDAPH3	ASINEX BAS00892957
ZINC01176821	Secin132	Consensus	Diversity	SVM GpiDAPH3	VitasM STK134679
ZINC02657221	Secin67	Cell adhesion	Similarity	MACCS	Enamine T5248258
ZINC07440926	Secin92	Cell adhesion	Similarity	GpiDAPH3	Enamine Z92738532
ZINC09223232	Secin105	Cell adhesion	Diversity	SVM GpiDAPH3	Enamine T5459575
ZINC09223915	Secin107	Cell adhesion	Diversity	SVM GpiDAPH3	Enamine T5463465
ZINC09984314	Secin114	Cell adhesion	Diversity	SVM GpiDAPH3	Enamine T5513572
ZINC08188318	Secin144	Cell adhesion	Similarity	GpiDAPH3	VitasM STK248770
ZINC09469141	Secin111	Cell adhesion	Diversity	SVM GpiDAPH3	Enamine T5649682
ZINC09469139	Secin110	Cell adhesion, exchange	Diversity	SVM GpiDAPH3	Enamine T5649690
ZINC02853425	Secin134	Cell adhesion, exchange	Diversity	SVM GpiDAPH3	VitasM STK133412
ZINC00693466	Secin2	Exchange	Similarity	GpiDAPH3	AMRI CGX-03198864
ZINC02074567	Secin23	Exchange	Diversity	SVM GpiDAPH3	ASINEX BAS02593302
ZINC03582297	Secin44	Exchange	Diversity	SVM GpiDAPH3	ChemDiv K415−0179
ZINC06496235	Secin46	Exchange	Diversity	SVM GpiDAPH3	ChemDiv K415−0299
ZINC00710394	Secin61	Exchange	Diversity	SVM GpiDAPH3	Enamine T0505−6944
ZINC00949959	Secin62	Exchange	Similarity	FP overlap	Enamine T0541−0562
ZINC07061995	Secin86	Exchange	Similarity	MACCS	Enamine T5424282
ZINC07158197	Secin87	Exchange	Similarity	MACCS	Enamine Z51130668
ZINC03582227	Secin122	Exchange	Diversity	SVM GpiDAPH3	LifeChemical F0546−0338
ZINC01206197	Secin133	Exchange	Diversity	SVM GpiDAPH3	VitasM STK233507
ZINC02890693	Secin69	FoxO, exchange	Diversity	SVM GpiDAPH3	Enamine T5891443
ZINC00693668	Secin3	FoxO	Similarity	FP overlap	AMRI CGX-03199266
ZINC04753173	Secin7	FoxO	Similarity	GpiDAPH3	AMRI CGX-02136485
ZINC06833057	Secin13	FoxO	Similarity	FP overlap	AMRI CGX-01218314
ZINC06872274	Secin15	FoxO	Similarity	FP overlap	AMRI CGX-01216930

[a]All compounds identified to be more active than SecinH3 are reported with their consecutive selection set number (ID), ZINC database code, and supplier information. [b]Search strategy (*i.e.*, similarity or diversity search). [c]Virtual screening method and molecular descriptor that identified each compound.

Virtual Screening Assessment. Of the 26 confirmed hits, 11 were identified by similarity and 15 by diversity searching (Table 1). Although all three FPs identified active compounds, there was a clear preference for the pharmacophore-type descriptor GpiDAPH3 in SVM calculations, where this FP identified all 15 hits (Table 1). None of the 26 active compounds was identified by both similarity and diversity searching, which highlights the need for complementary methodologies in computational hit identification (*2*). Interestingly, the most active compounds in individual assays were identified by SVM calculations. However, compound potency was not utilized as a parameter in any of these VS calculations (which generally applies to current VS approaches). We have applied different standard VS methods using a search protocol specifically tailored toward available Secin reference compound information. Hence, we would anticipate that other ligand-based VS approaches would in this case also be capable of identifying active compounds.

Structural Diversity. The 26 inhibitors were characterized by a high degree of scaffold (core structure) diversity. The spectrum of structurally diverse active compounds is illustrated in Figure 4. A total of 23 different chemical scaffolds were observed among active compounds that corresponded to 21 unique carbon skel-

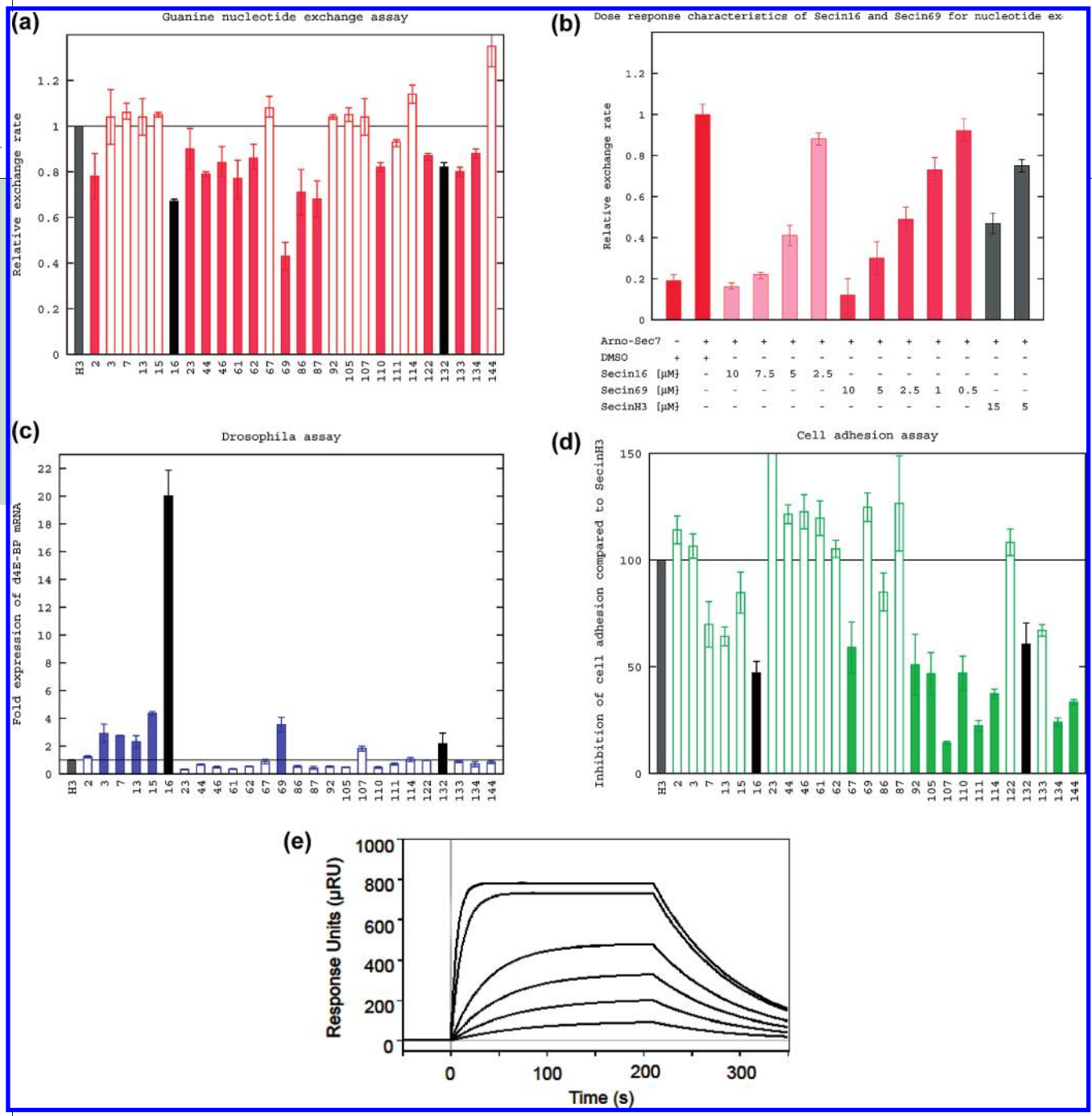

Figure 2. Experimental results. Assay results are shown for all tested compounds that were more active than SecinH3 in at least one of three assay systems. Threshold lines indicate the inhibition level of SecinH3. Filled bars represent inhibitors that were more active than SecinH3 according to our selection criteria. Results for the two consensus hits (Secin16 and 132) are shown in black. a) Guanine nucleotide exchange. b) Dose–response characteristics of Secin16 and Secin69 in nucleotide exchange. c) *Drosophila* assays. d) Cell adhesion assay; the average of PMA- and OKT3-based adhesion is reported. e) Surface plasmon resonance sensorgrams of the interaction of Secin16 with the immobilized Sec-7 domain of cytohesin-1 at the indicated concentrations. Shown are fitted binding and dissociation curves. From the top to the bottom, tracings correspond to the following concentrations of Secin16: 100, 50, 10, 5, 2.5, and 1 μM.

etons (see Methods) distinct from that of SecinH3. Active molecules sharing the SecinH3 core structure were not identified. Furthermore, only two subsets of active compounds had the same carbon skeleton: Secin16 and 132, and Secin3, 13, 15, 110, and 111, respectively. The compounds that were most active in each assay had distinct structures, as illustrated in Figure 3.

The high degree of structural diversity among the inhibitors raises the question why holistic molecular similarity methods were successful in this case. Detailed structural comparisons of inhibitors provided some clues. In Figure 4, two overlapping molecular regions in SecinH3 are highlighted, each consisting of three rings. The substructure highlighted in pink contains the tria-

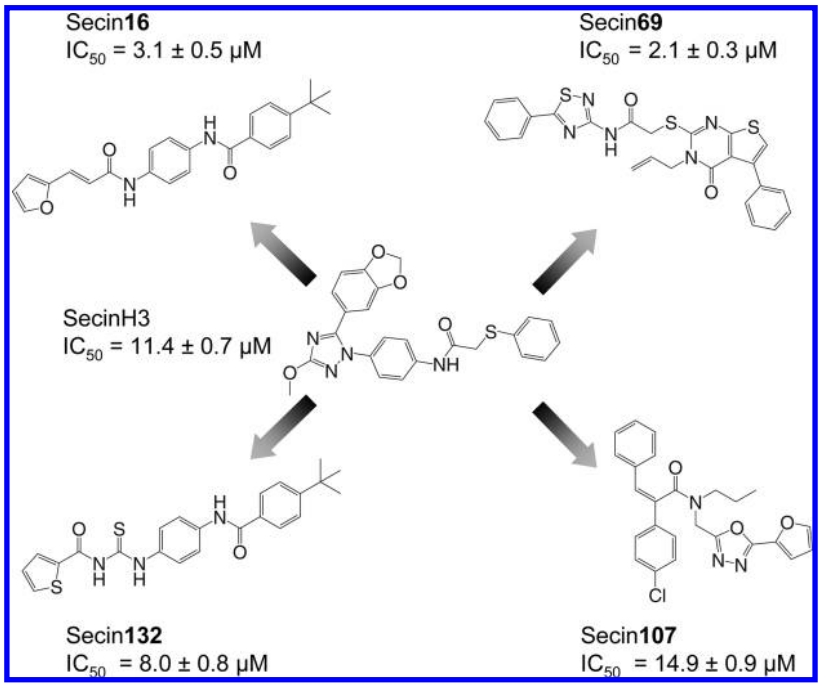

Figure 3. Secin comparison. The structures of SecinH3, 16, 69, 107, and 132 are shown, and IC_{50} values for nucleotide exchange inhibition are reported.

zole moiety in the center with ring substituents at its 1- and 2-position, and the cyan substructure contains the 2-substitued triazole as the terminal ring. Although the inhibitors were characterized by a high degree of structural diversity, substructures resembling one or the other of these SecinH3 fragments were recurrent among them. For example, all inhibitors of the inner compound diversity layer in Figure 4 contained substructures reminiscent of the pink SecinH3 fragment, and most inhibitors in the outer two layers contained substructures reminiscent of either the pink or cyan fragment. These substructure resemblances explain why similarity methods were able to detect an array of structurally diverse active compounds by extrapolating from the SecinH3 chemotype.

Differentiated Functions. In addition to structural diversity, active compounds were characterized by differential inhibitory profiles (Table 1, Figure 4). Two inhibitors sharing the same scaffold, Secin16 and Secin132, were strongly active in all three assays. Other compounds were more active than SecinH3 in two of three or individual assays. One compound had dual activity in inhibiting *Drosophila* insulin signaling and nucleotide exchange (Secin69) and two others had dual activity in cell adhesion and nucleotide exchange assays (Secin110 and 134). By contrast, no compound with dual activity in *Drosophila* insulin signaling and cell adhesion assays was identified. Furthermore, four compounds were predominantly active in down-regulating *Drosophila* insulin signaling, 10 in nucleotide exchange, and seven in cell adhesion assays. Thus, chemotype diversity corresponded to a spectrum of differentiated inhibitory functions. However, subtle structural changes among active compounds were also sufficient to act as a functional switch, as revealed by the two compound subsets sharing very similar core structures, as discussed above. For example, Secin13 was only active in inhibiting insulin signaling, whereas Secin110 sharing the same core structure was active in cell adhesion and GDP/GTP exchange assays, but not in inhibiting insulin signaling. Moreover, three compounds within this series were only active in down-regulating insulin signaling (Secin3, 13, and 15), another active compound exclusively abrogated cell adhesion (Secin111), and the remaining compound inhibited cell adhesion and nucleotide exchange (Secin110). Among compounds with differential inhibitory profiles, four strong cell adhesion inhibitors were identified (Secin107, 111, 114, and 144), which were either comparable to SecinH3 in nucleotide exchange assays (Secin107 and 111) or less active (Secin114 and 144). Thus, although these compounds inhibited cytohesin-dependent nucleotide exchange, they were predominantly inhibitors of signaling events leading to adhesion, in contrast to SecinH3.

Conclusions. In our study, we have investigated the potential of VS approaches to identify novel active compounds as molecular probes for studying multifunctional cytohesin proteins. For this target protein family, the identification of novel inhibitors was also highly rel-

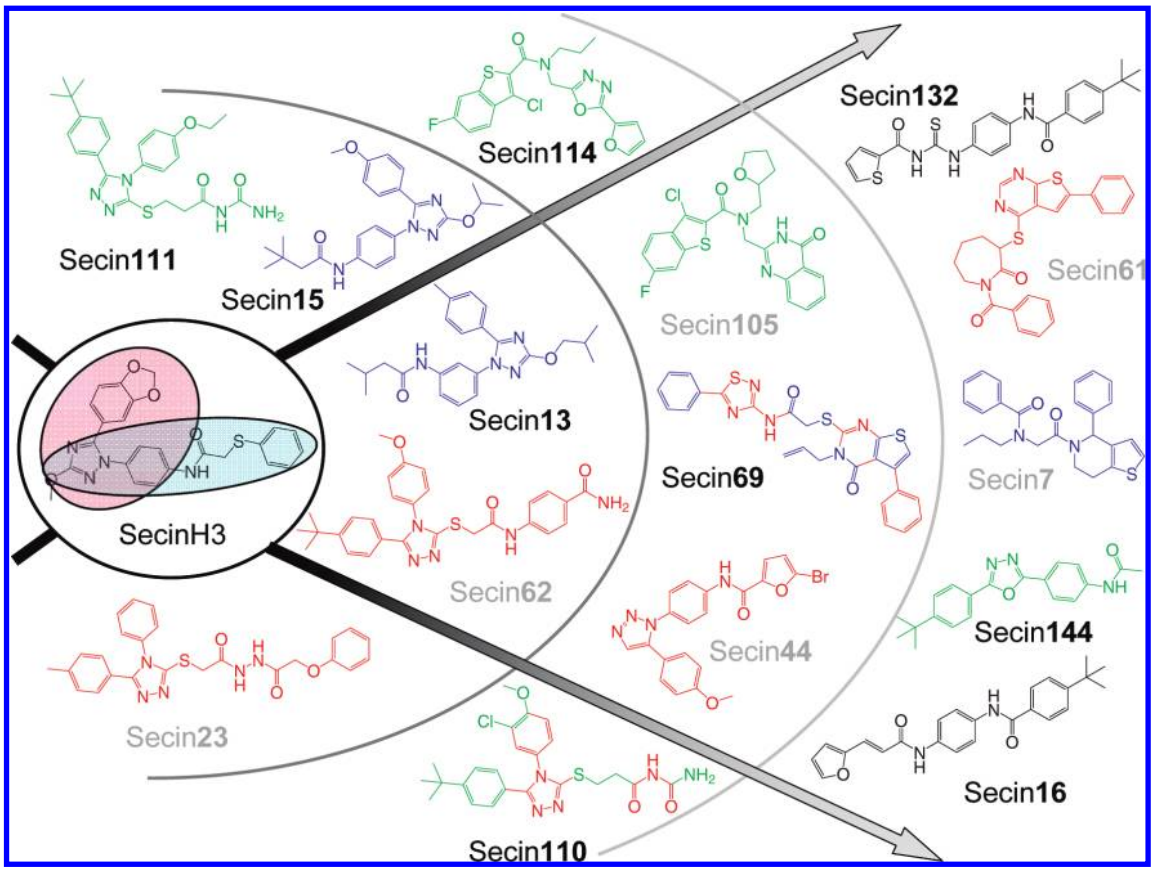

Figure 4. Structural spectrum of active compounds. From the left to the right, exemplary hits from all assays are shown and arranged in layers that are structurally increasingly dissimilar to SecinH3. In SecinH3, two overlapping molecular regions are highlighted, each consisting of three rings (shown in pink and cyan, respectively), which resemble alternative substructures found in many inhibitors. The color scheme reflects assay activity: red, GDP/GTP exchange; blue, *Drosophila* assays; green, cell adhesion assays. Compounds active in two assays are bicolored and compounds active in all three assays are shown in black. Molecules discussed in the text have black labels.

evant for practical applications. Homologous genes often exhibit redundant or promiscuous physiological activities, making it difficult to study their function by genetic knock-out or siRNA knock-down. Unambiguous phenotypic effects are frequently observed only when two or more members of a homologous protein family are targeted simultaneously. Pharmacological approaches in which *pan*-target family inhibitors are applied have proven a useful alternative to these genetic strategies. Accordingly, with SecinH3, a *pan*-cytohesin inhibitor was successfully utilized to elucidate the involvement of cytohesins and Steppke in insulin signaling. By contrast, SecinH3 was only a weak inhibitor of signaling events leading to cytohesin-mediated cell adhesion. Thus, the SecinH3 inhibitory profile provided a meaningful starting point for our VS analysis.

A key question has been whether computational extrapolation from SecinH3 might identify compounds that not only are active but display differential inhibitory profiles, an issue that had not been addressed before. By evaluating VS candidate compounds selected in three different assay systems, an array of structurally diverse active molecules was indeed identified that inhibited cytohesin functions with highly differentiated inhibitory profiles. It cannot be ruled out that newly identified active compounds also bind to targets other than cytohesins and thereby elicit differential functional effects. However, the template compound SecinH3 is known to

inhibit nucleotide exchange, insulin signaling, and cell adhesion (albeit only weakly), and the new inhibitors are characterized by the same functional profile but display changes in the relative magnitude of inhibitory effects. Because all new inhibitors specifically bind to the Sec-7 domain, it is likely that this domain is their primary target (but again, binding of these compounds and also SecinH3 to additional targets can not be ruled out).

A total of 26 newly identified inhibitors were more potent than SecinH3 in one or more assays, which is unusual for active compounds identified by VS, including strong inhibitors of cytohesin-mediated cell adhesion. The results indicate that VS approaches have the potential to identify molecular probes with differentiated functions, despite intrinsic limitations of computational screening. Methodologically, it is interesting to note that SVM calculations utilizing 2D molecular representations and only a few active training examples were very effective in detecting compounds with Secin activity. Considering general VS standards (*1–3*), the search for cytohesin Sec-7 domain-directed inhibitors has been a perhaps surprisingly successful project, in terms of the number of novel active compounds that were identified, their potency relative to that of the reference molecules, and their degree of structural diversity. Different targets display different ligand binding constraints, and success rates of compound screening often vary considerably. Varying degrees of substructure resemblance between different inhibitors that we have observed here, as discussed above, might suggest that binding of Secins to the Sec-7 domain can be achieved through overlapping alternative pharmacophore elements. Because all inhibitors were active in nucleotide exchange assays, they are very likely to bind to the same site, but variations in binding modes might perhaps occur. Whether different binding modes might exist that correspond to different functions or multiple binding sites remains to be elucidated. Regardless, the newly identified cytohesin inhibitors represent a spectrum of structural and functional diversity. Secin16 and 132 consistently improved the efficacy of SecinH3 in all three assay systems and should hence be attractive molecular probes for detailed investigations of cytohesin signaling functions. Moreover, other compounds, in particular Secin107 and 144, were found to be strong inhibitors of cytohesin-mediated cell adhesion, thus further extending the potential to analyze different cytohesin functions using small molecules.

METHODS

Virtual Screening. For similarity searching, three 2D molecular fingerprints were utilized: MACCS (*14*), Molprint2D (*15*), and GpiDAPH3 (Molecular Operating Environment, MOE) (*16*). These 2D FPs transform 2D molecular graphs into vectors of binary values. MACCS consists of 166 catalogued structural fragments whose presence or absence in a test molecule is monitored. By contrast, Molprint2D codes for the presence of layered atom environments (*15*). Furthermore, GpiDAPH3 is a 3-point pharmacophore-type FP that assigns to each atom in a molecule one of eight atom types computed from three atomic properties (*i.e.*, "pi system", "donor", and "acceptor"). MACCS and GpiDAPH3 were calculated using the MOE (http://www.chemcomp.com) and Molprint2D using public domain software (http://www.molprint.com). Fingerprint overlap between reference and database compounds was quantified as a measure of molecular similarity using the Tanimoto coefficient (Tc) (*17*) and database compounds were ranked by decreasing similarity values.

For support vector machine calculations, a publicly available implementation was applied (SVM^{light}) (*18*), and two FPs, Molprint2D and GpiDAPH3, were used as descriptors, respectively. SVMs (*19–21*) are algorithms for supervised machine learning that were originally developed for binary object classification. SVM learning takes active (positive training examples) and inactive molecules (negative training examples) into account. The basic idea of SVM learning is to construct a hyperplane (*H*) in descriptor space that best separates positive and negative training examples by minimizing the classification error and maximizing the generalization performance to avoid overfitting. The resulting model is then applied to classify objects (database compounds) as active or inactive depending on which side of *H* they fall. Moreover, in order to obtain a ranking, database compounds are sorted according to their distance from *H*. Hence, compounds located in the positive half-space are ranked by decreasing distances from *H*, followed by compounds in the negative half-space ranked by increasing distances from *H*.

Computational screening was applied to ~2.6 million virtually formatted publicly available ZINC compounds (*22*) that were originally collected from commercial sources. FP database rankings were filtered to remove compounds having a MACCS Tc value greater than 0.95 compared to SecinH3 (*i.e.*, nearly identical compounds). SVM rankings were filtered to retain only molecules with a MACCS Tc smaller than 0.85 to any known active compound, hence supporting the selection of structurally diverse candidates.

Molecular Scaffold Analysis. Scaffolds were isolated from active compounds by removal of all R-groups from ring systems and aliphatic moieties linking ring systems. In addition, scaffolds were transformed into carbon skeletons by replacing all heteroatoms with carbons and setting all bond orders to 1 (*i.e.*, single bonds).

Guanine Nucleotide Exchange Assay. A total of 145 test compounds were acquired from commercial sources according to Table 1. Purity of these compounds was confirmed on the ba-

sis of NMR and MS data provided by vendors. [Δ17]ARF1 and ARNO-Sec-7 were subcloned into pET15 vectors (Novagen) as described previously (*12, 23–25*). N-Terminal truncated [Δ17]ARF1 (amino acids 18−181), lacking the first 17 amino acids and ARNO-Sec-7 (amino acids 50−255 of ARNO) were expressed in *Escherichia coli* and purified by Ni-NTA chromatography (Ni-NTA agarose, Quiagen). GDP/GTP exchange was measured on [Δ17]ARF1 by tryptophan fluorescence because a large increase in intrinsic fluorescence of ARF occurs upon exchange of GDP for GTP (*25, 26*). All measurements were performed in PBS pH 7.4, 3 mM MgCl2 at 37 °C. [Δ17]ARF1 (1 μM) in PBS without $MgCl_2$ was preincubated with GDP (80 μM) in the presence of EDTA (2 mM) for 15 min. The bound GDP was stabilized by addition of $MgCl_2$ (final concentration 3 mM) and incubation for 5 min. For each exchange reaction 250 nM [Δ17]ARF1 was mixed with 10 nM ARNO-Sec-7 (total volume 200 μL) in the absence or presence of inhibitor. The reaction was started by injection of GTP (50 μM). The tryptophan fluorescence was measured at excitation and emission wavelength of 280 and 340 nm, respectively. All fluorescent measurements were performed with a Varioskan microplate reader (Thermo Scientific), in 96-well plates. For analysis all data were fitted by linear regression. IC_{50} values were determined in 5-fold repeated titration assays.

***Drosophila* Assays.** In order to evaluate potential interference of candidate molecules with *Drosophila* insulin signaling, S2 cells (Schneider 2 insect cell line, Invitrogen) were grown under starvation conditions and subsequently treated with insulin. This triggers a strong activation of the insulin signaling cascade resulting in the nuclear exclusion of dFOXO, which in turn down-regulates the transcriptional repressor d4E-BP. In contrast, when compounds inhibit the GEF activity of Steppke insulin signaling is blocked and nuclear dFOXO activity increases, leading to an up-regulation of d4E-BP transcription. This is measured by quantitative RT-PCR and compared to the SecinH3 effect in the same assay.

Cell Culture and Compound Treatment. S2 cells (0.5×10^6) were grown in 24-well dishes in medium without FCS representing starvation condition. Six hours after seeding cells were treated for 18 h with 10 μM compound or 10 μM SecinH3 as reference. Twenty-two hours after seeding cells were treated with insulin (5 μg mL^{-1}, Sigma). Controls were performed without compound and/or without insulin treatment. Compounds were dissolved in DMSO. Final DMSO concentration in the cell culture medium was 0.5%.

RNA Isolation, cDNA synthesis and Real-Time PCR. A NucleoSpin 8 RNA Kit (Macherey-Nagel) was used for RNA preparation. Genomic DNA digestion and first strand cDNA synthesis was carried out with 250 ng of total RNA using QuantiTect Reverse Transkription Kit (Qiagen). Quantitive PCR was performed with the iQ5 Real-Time PCR detection system (Bio-Rad) and iQ SYBR Green Super Mix (Bio-Rad). Three reactions were done in parallel for each template. Real time PCR was analyzed using Bio-Rad iQ5 Optical System Software and Microsoft Excel.

Cell Adhesion Assays. The ability of candidate compounds was tested to inhibit adhesion of the T cell line Jurkat E6 to an immobilized ICAM-1-Fc-fusion protein. On the T cell side, this interaction is exclusively mediated by the β-2 integrin LFA-1 that was enabled to bind to immobilized ICAM-1 through stimulation of Jurkat cells with an anti-TCR antibody or with the phorbol ester PMA (*9*). Cell adhesion to ICAM-Fc was carried out on 96-well dishes and was read out by fluorescence detection of total input verus bound cells which had been labeled with the DNA dye H33342.

Jurkat E6.1 cells were incubated in RPMI, 0.5% DMSO, and where indicated, 25 μM of the respective compound for 1 h at 37 °C, 5% CO_2. After 30 min the cells were labeled with the fluorochrome bisbenzimide trihydrochloride H33342 (Calbiochem) at 12 mg mL^{-1} for 30 min at 37 °C. Two $\times 10^5$ cells/well were subsequently dispensed into a 96-well plate at 100 μL $well^{-1}$. Prior to use plates were coated with 12 μg mL^{-1} goat antihuman IgG for 90 min at 21 °C, blocked with 1% BSA in PBS overnight and incubated for 60 min with culture supernatants from CV-1 cells expressing an ICAM-1-Fc fusion protein.

Cells were either stimulated with 5 μg mL^{-1} purified anti-CD3 antibody (OKT3) or with PMA (50 ng mL^{-1}), respectively. Cells were allowed to adhere to the ICAM-1-Fc coated dishes for 15 min and unbound cells were washed off with HBSS. Adherent cells were read out using a fluorescence plate reader (Synergy-HT1, Biotek) at 485 nm in triplicates. "100% adhesion" corresponds to the mean of OKT3 and PMA stimulated samples treated with 25 μM SecinH3 as reference.

Surface Plasmon Resonance Assays. SPR experiments were performed using a dual-channel SR7000DC system (Reichert Inc.). Recombinant cytohesin-1 Sec-7 domain was covalently immobilized on a HC1500m chip (Xantec bioanalytics GmbH). The surface was activated with activation buffer (0.1 M NHS, 0.7% EDC, 0.05 M MES, pH 5.0), and the protein was applied at a concentration of 50 μg mL^{-1} in 5 mM acetic acid, pH 4.5 to the sample channel only. Unreacted residues on the chip surface were quenched with 1 M ethanolamine, pH 8.5. Binding and dissociation were performed in 1% DMSO containing PBS at a flow rate of 50 μL min^{-1}. Regeneration of the chip surface was achieved by injection of 10 mM glycine HCl, pH 3.0. The net sample channel response (which was calculated by subtracting the response of the reference channel from that of the sample channel) was corrected for blank buffer injection and DMSO injection. Each curve represents at least three experiments. Data processing and curve fitting was done using SPR V4.0.17 (Reichert Inc.) and Scrubber2 software (BioLogic Software).

Acknowledgment: We thank B. von der Gönna for her help with compound acquisition, K. Rotscheidt for preparing the compounds for screening, and A. Hall for help with the *Drosophila* S2 experiments. The work has been supported by *Sonderforschungsbereich 704* of the *Deutsche Forschungsgemeinschaft*.

REFERENCES

1. Shoichet, B. K. (2004) Virtual screening of chemical libraries, *Nature 432*, 862–865.
2. Bajorath, J. (2002) Integration of virtual and high-throughput screening, *Nat. Rev. Drug Discovery 1*, 882–894.
3. Eckert, H., and Bajorath, J. (2007) Molecular similarity analysis in virtual screening: foundations, limitations, and novel approaches, *Drug Discovery Today 12*, 225–233.
4. Naylor, E., Arredouani, A., Vasudevan, S. R., Lewis, A. M., Parkesh, R., Mizote, A., Rosen, D., Thomas, J. M., Izumi, M., Ganesan, A., Galione, A., and Churchill, G. C. (2009) Identification of a chemical probe for NAADP by virtual screening, *Nat. Chem. Biol. 5*, 220–226.
5. Klarlund, J. K., Guilherme, A., Holik, J. J., Virbasius, J. V., Chawla, A., and Czech, M. P. (1997) Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec-7 homology domains, *Science 275*, 1927–1930.
6. Fuss, B., Becker, T., Zinke, I., and Hoch, M. (2006) The cytohesin Steppke is essential for insulin signaling in Drosophila, *Nature 444*, 945–948.
7. Ogasawara, M., Kim, S. C., Adamik, R., Togawa, A., Ferrans, V. J., Takeda, K., Kirby, M., Moss, J., and Vaughan, M. (2000) Similarities in function and gene structure of cytohesin-4 and cytohesin-1, guanine nucleotide-exchange proteins for ADP-ribosylation factors, *J. Biol. Chem. 275*, 3221–3230.
8. Kolanus, W., Nagel, W., Schiller, B., Zeitlmann, L., Godar, S., Stockinger, H., and Seed, B. (1996) αLβ2 integrin/LFA-1 binding to ICAM-1 induced by cytohesin-1, a cytoplasmic regulatory molecule, *Cell 86*, 233–242.

9. Ley, K., Laudanna, C., Cybulsky, M. I., and Nourshargh, S. (2007) Getting to the site of imflammation: the leukocyte adhesion casacade updated, *Nat. Rev. Immunol. 7*, 678–689.
10. Kliche, W., Nagel, W., Kremmer, E., Atzler, C., Ege, A., Knorr, T., Koszinowski, U., Kolanus, W., and Haas, J. (2001) Signaling by human herpesvirus 8 kaposin A through direct membrane recruitment of cytohesin-1, *Mol. Cell 7*, 833–843.
11. Perez, O. D., Mitchell, D., Jager, G. C., South, S., Murriel, C., McBride, J., Herzenberg, L. A., Kinoshita, S., and Nolan, G. P. (2003) Leukocyte functional antigen 1 lowers T cell activiation thresholds and signaling through cytohesin-1 and Jun-activating binding protein 1, *Nat. Immunol. 4*, 1083–1092.
12. Hafner, M., Schmitz, A., Grüne, I., Srivatsan, S. G., Paul, B., Kolanus, W., Quast, T., Kremmer, E., Bauer, I., and Famulok, M. (2006) Inhibition of cytohesins by SecinH3 leads to hepatic insulin resistance, *Nature 444*, 941–944.
13. Geiger, C., Nagel, W., Boehm, T., van Kooyk, Y., Figdor, C. G., Kremmer, E., Hogg, N., Zeitlmann, L., Dierks, H., Weber, K. S., and Kolanus, W. (2000) Cytohesin-1 regulates beta-2 integrin-mediated adhesion through both ARF-GEF function and interaction with LFA-1, *EMBO J. 19*, 2525–2536.
14. http://www.mdl.com
15. Bender, A., Mussa, H. J., Glen, R. C., and Reiling, S. (2004) Similarity searching of chemical databases using atom environment descriptors (MOLPRINT2D): evaluation of performance, *J. Chem. Inf. Comput. Sci. 44*, 1708–1718.
16. http://www.chemcomp.com
17. Willett, P. (2005) Searching techniques for databases of two- and three-dimensional structures, *J. Med. Chem. 48*, 4183–4199.
18. Joachims, T. (1999) *Making Large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning*(Schölkopf, B.Burges, C.Smola, A., Eds.) Chapter 11, MIT Press, Cambridge, MA.
19. Burges, C. J. C. (1998) A tutorial on support vector machines for pattern recognition, in *Data Mining and Knowledge Discovery*, Vol. 2, pp 121–167, Kluwer Academic Publishers, Boston.
20. Vapnik, V. N. (2000) *The Nature of Statistical Learning Theory*, 2nd ed., Springer, New York.
21. Boser, B. E., Guyon, I. M.Vapnik, V. (1992) *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp 144–152, Pittsburgh, PA, 1992, ACM, New York.
22. Irwin, J. J., and Shoichet, B. K. (2005) ZINC - a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model. 45*, 177–182.
23. Hafner, M., Vianini, E., Albertoni, B., Marchetti, L., Grüne, I., Gloeckner, C., and Famulok, M. (2008) Displacement of protein-bound aptamers with small molecules screened by fluorescence polarization, *Nat. Protoc. 3*, 579–587.
24. Bi, X., Schmitz, A., Hayallah, A. M., Song, J. N., and Famulok, M. (2008) Affinity-based labeling of cytohesins with a bifunctional SecinH3 photoaffinity probe, *Angew. Chem., Int. Ed. 47*, 9565–9568.
25. Antonny, B., Beraud-Dufour, S., Chardin, P., and Chabre, M. (1997) N-terminal hydrophobic residues of the G-protein ADP-ribosylation factor-1 insert into membrane phospholipids upon GDP to GTP exchange, *Biochemistry 36*, 4675–4684.
26. Northup, J. K., Smigel, M. D., and Gilman, A. G. (1982) The guanine nucleotide activating site of the regulatory component of adenylate cyclase. Identification by ligand binding, *J. Biol. Chem. 257*, 11416–11423.
27. Baaske, P., Wienken, Christoph, J., Reineck, P., Duhr, S., and Braun, D. (2010) Optical thermophoresis for quantifying the buffer dependence of aptamer binding, *Angew. Chem., Int. Ed. 49*, 2238–2241.
28. Duhr, S., and Braun, D. (2006) Why molecules move along a temperature gradient, *Proc. Natl. Acad. Sci. U.S.A. 103*, 19678–19682.