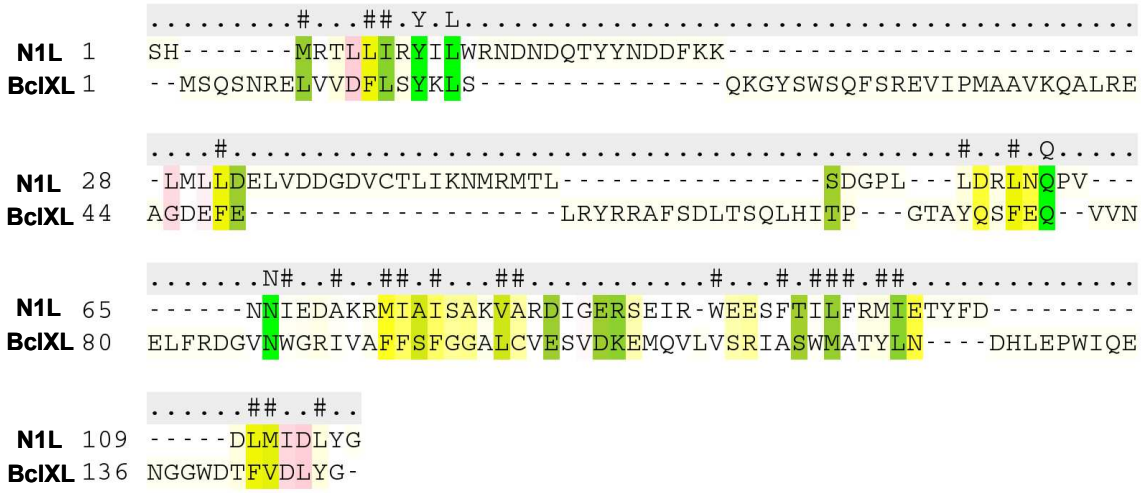
Supporting Information.

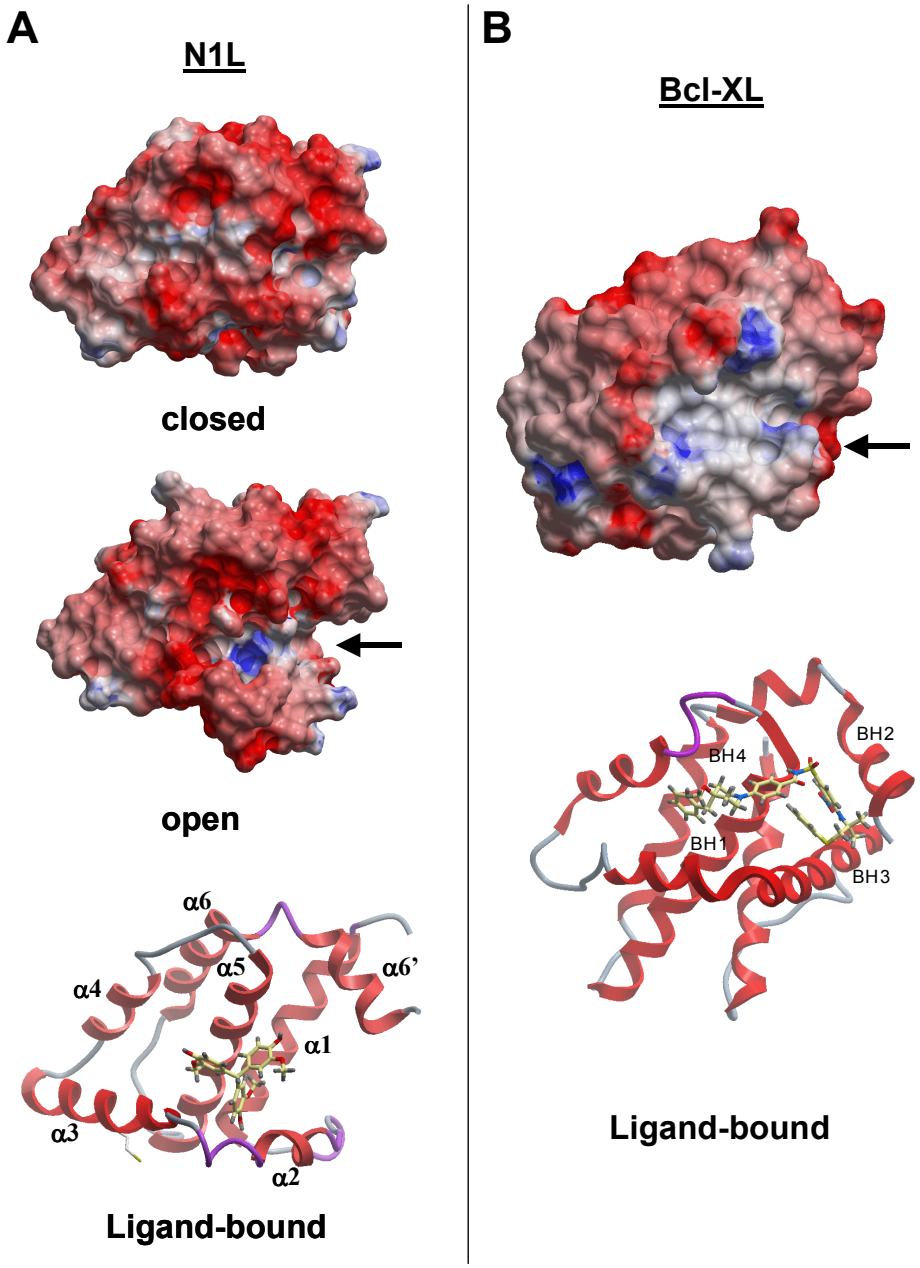
Vaccinia Virus Virulence Factor N1L is a Novel Promising Target for Antiviral Therapeutic Intervention

Anton V. Cheltsov[1,2], Mika Aoyagi[1], Alexandre Aleshin[1], Yu Eric Chi-Wang[1], Taylor Gilliland[1], Dayong Zhai[1], Andrey A. Bobkov[1], John C. Reed[1], Robert C. Liddington[1*] and Ruben Abagyan[2*]

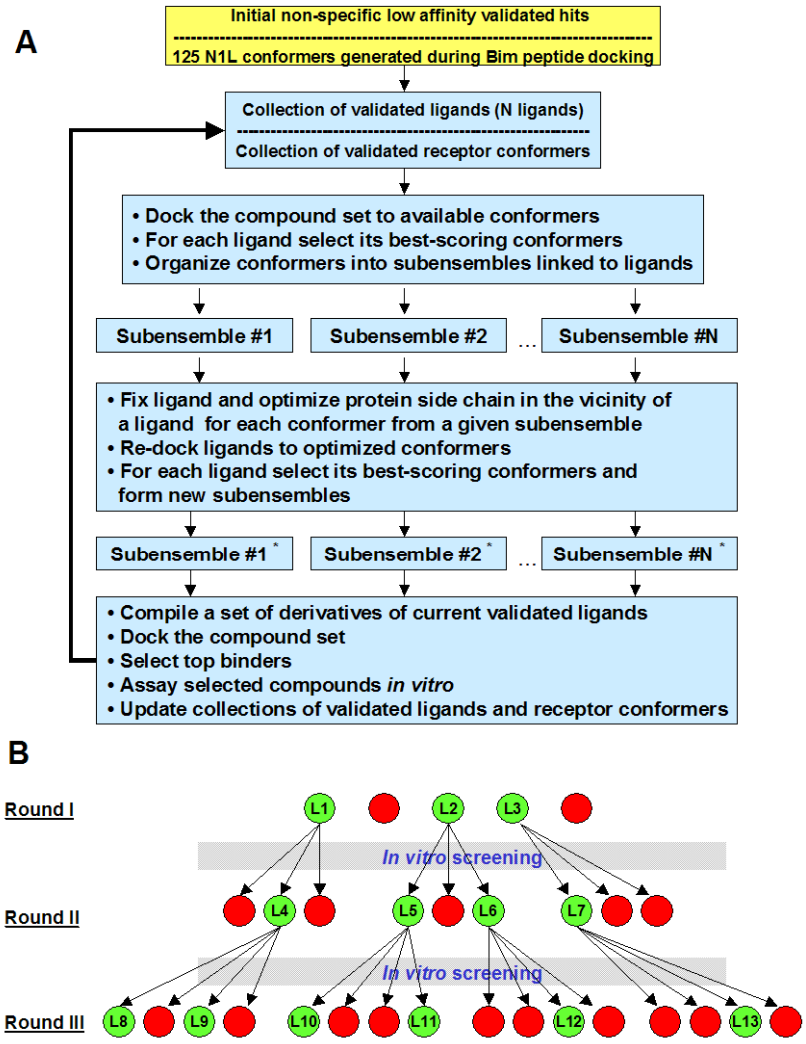
[1]Infectious and Inflammatory Disease Center, Burnham Institute for Medical Research, La Jolla, CA 92037, USA; [2]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA



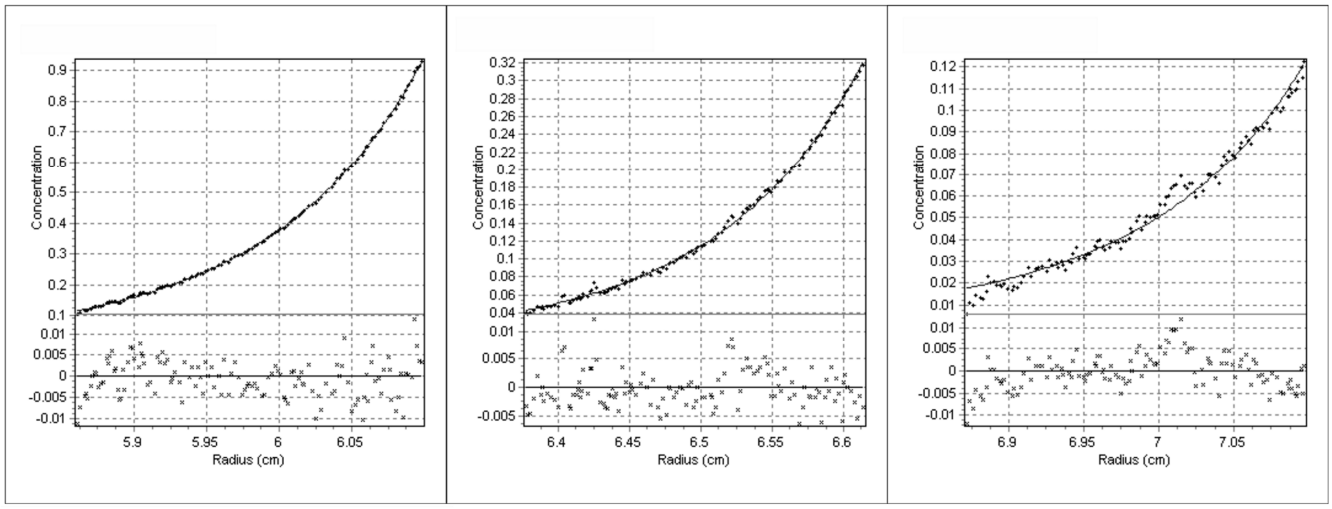
SUPPLEMENTARY FIGURE 1. **The amino acid sequence alignment of *vaccinia* N1L protein and Bcl-XL.** Sequence identity = 3%.



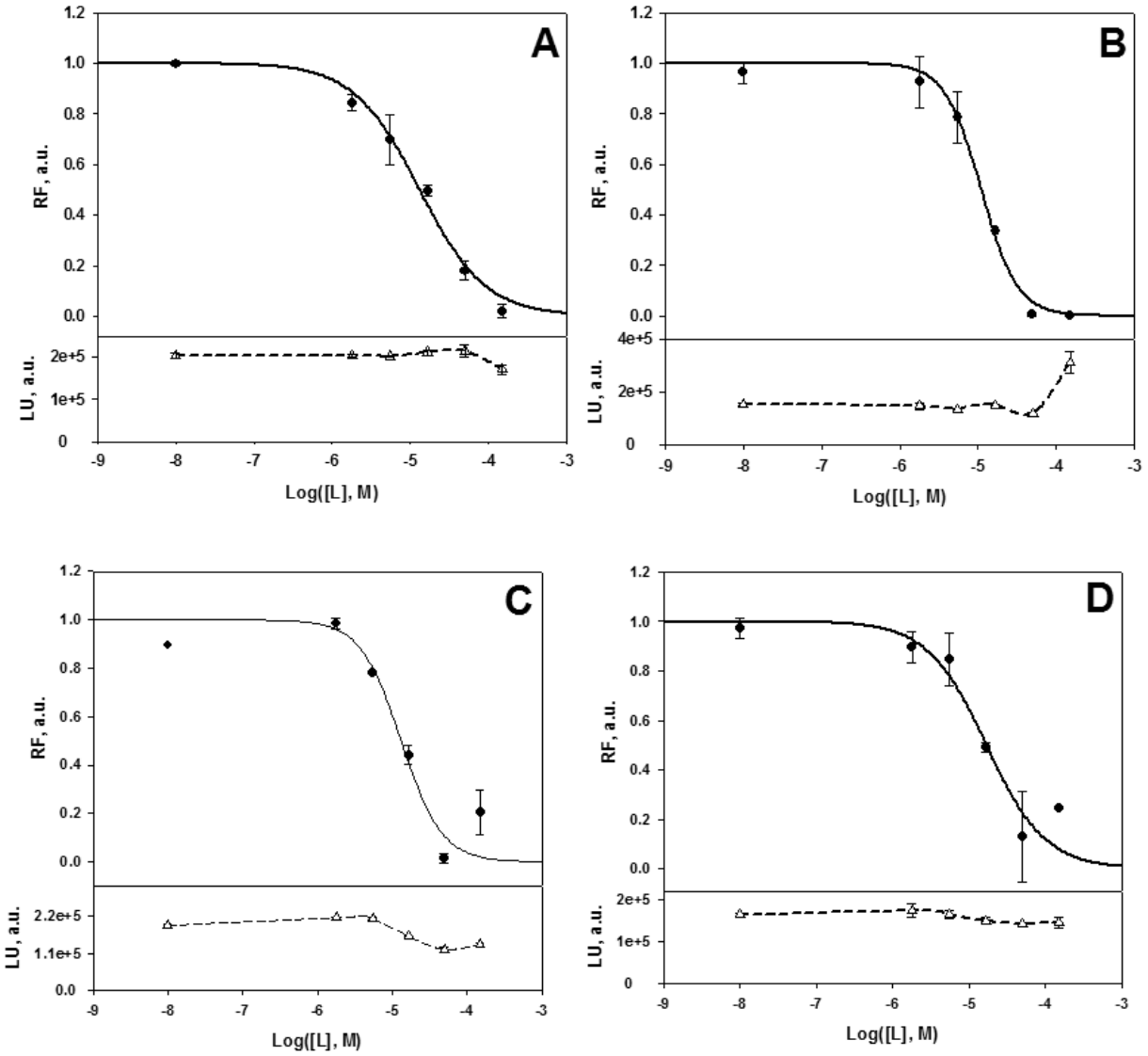
SUPPLEMENTARY FIGURE 2. **Modeling of N1L induced-fit-like state. A.** Native, "closed" N1L monomer structure, and a model of "open" N1L monomer after docking to Bim BH3 peptide (see *Methods*). Arrow indicates the location of the cavity chosen as a site for *in silico* ligand docking experiments. A representative low energy conformer is shown. The probable binding mode of a ligand is shown in cartoon style. N1L helices are labeled. **B.** The crystal structure of Bcl-XL (PDB# 1R2D) is shown as a molecular surface mesh. The location of BH3 domain is indicated by arrow. The complex with an acyl-sulfonamide-based ligand is shown at the bottom as a cartoon (NMR, PDB# 2O2M). Important BH regions are indicated on a top of the structure. Molecular surfaces were colored by partial atomic charges, corrected for interactions with the solvent.



SUPPLEMENTARY FIGURE 3. **Strategy for optimization of both receptor induced-fit-like states and ligand scaffolds. A. Block-diagram of a computational strategy applied to the discovery and optimization of the N1L antagonists.** Given initial the sets of N1L conformers and validated low affinity non-specific ligands, the procedure tries to optimize both, based on biological properties of the compounds. The number of both conformers and active ligands can vary from iteration to iteration. The procedure is repeated iteratively until functional convergence is achieved (see *Methods* for details). **B. The N1L receptor and ligand optimization logistics.** The N1L conformer subensembles are shown as circles. The validated and discarded conformer subensembles are colored green and red, respectively. The validated conformers were used as seeds for the generation of new subensembles of N1L conformers (indicated by arrows). The validated ligands are inside the circles, corresponding to the subensembles where the best docking score was achieved. The ligands labels are shown for clarity only; they do not correspond to ligand numbering as in the Supplementary Table 1. See *Methods* for details.



SUPPLEMENTARY FIGURE 4. **Sedimentation equilibrium of N1L in the presence of ligand 11.** Data shown for three concentrations used: 0.45, 0.15 and 0.05 mg/ml. Solid lines show the best-fit achieved using monomer-dimer model (K_D^{11} = 8 μM). Plots of the residuals for each data set are shown bellow.



SUPPLEMENTARY FIGURE 5. **Antiviral properties of select N1L antagonists. A. 9**, EC_{50} = 13.2 μM, CV1 cells; **B. 9**, EC_{50} = 10.9 μM, HT1080 cells; **C. 11**, EC_{50} = 13.2 μM, CV1 cells; **D. 11**, EC_{50} = 16.9 μM, HT1080 cells. RF, normalized GFP fluorescence; LU, luminescence (absolute values); Log([L], M), decimal logarithm of ligand concentration in M. Data were fit to the variable slope dose response equation. Refer to Table 1 for ligand structures. See *Methods* for assay conditions.

SUPPLEMENTARY TABLE 1. **Summary of ligands optimization.**

Ligand ID	NCI NSC#	IC_{50}, μM	
		[N1L] = 1 μM	[N1L] = 0.5 μM
ROUND I			
1	16953	35.0	-
2	**19939**	**12.4**	**8.8**
3	53268	46.0	-
4	53272	54.0	-
5	53274	39.0	-
6	53276	16.0	-
7	156573	31.0	-
8	273403	55.0	-
9	**369686**	**6.6**	**5.6**
10	**638495**	**7.5**	**5.0**
11	Resveratrol*	21.9	-
ROUND II			
12	**31761**	**17.9**	**-**
13	**155590**	**3.5**	**2.4**
14	**163797**	**15.7**	**6.3**
15	**401220**	**2.0**	**1.8**
16	631366	28.9	19.5
17	**648419**	**20.5**	**-**
ROUND III			
18	7388	13.7	-
19	17061	6.1	-
20	26679	26.5	-
21	34688	26.2	-
22	36798	7.8	-
23	37214	1.9	0.6
24	52934	13.7	-
25	112806	25.9	-
26	126395	3.6	1.5
27	136955	5.7	-
28	150537	16.3	-
29	153166	1.3	1.2
30	155497	1.7	-
31	170006	1.6	0.9
32	209920	9.3	-
33	269124	18.5	-
34	348718	5.0	-
35	348905	13.9	-
36	402887	36.4	-
37	408014	31.4	-
38	607391	46.1	-
39	647364	48.8	-
40	647369	44.8	-
41	664154	1.8	0.9

Highlighted rows in bold correspond to ligands, chosen for optimization in subsequent protocol rounds. For Round III, the hits from both Round I and II were chosen. IC_{50}s were determined as described in *Methods*. For more potent ligands, N1L concentration of 0.5 μM was used in competitive binding experiments to obtain better IC_{50} values (IC_{50} asymptotically approaches true K_D with the decrease of a receptor concentration). *Resveratrol was added to the compound list because of its structural similarity to ligand **3** (ID10) (Table 1).

Supplementary Materials and Methods

Discovery of Initial N1L Inhibitor Hits. The initial hits were discovered by using standard virtual ligand screening (VLS) protocol as implemented in the ICM program[1]. The VLS experiments were performed with each of 125 receptor conformers, retained after Bim peptide docking. The ligand database, used for *in silico* docking, was constructed from complete NCI database by clustering it by chemical similarity (as implemented in the ICM program). The final database contained approximately 10000 diverse ligands. For each N1L conformer, top predicted binders were retained (the ICM docking score cutoff was applied). The hit lists for each conformer were than pooled together. The hit list was then sorted by the docking score and redundant ligand entries were eliminated. This procedure produced the intersection of the top binders across 125 N1L conformers. The set of top predicted binders was further clustered by chemical similarity, and cluster representatives were ordered from NCI for *in vitro* screening.

In Vitro Driven In Silico Optimization of Both Receptor and Ligands. The basic idea of the ligand optimization protocol is schematically represented in the Supplementary Figure 3. At each round of the protocol two information pools are maintained: a collection of validated ligands and a collection of validated receptor conformers organized into subensembles. Within the protocol framework a validated ligand is defined as a compound, which passed a battery of independent *in vitro* tests. A validated conformer subensemble is defined as a set of receptor conformers where a validated ligand achieved the best docking score with at least one of the conformers. These information pools are updated with refined ligand scaffolds and receptor conformers during iterative application of the protocol. The receptor conformers can persist from iteration to iteration, while only derivatives of validated ligands are passed to the next iteration.

The progression from one iteration of the protocol to another is as follows. First, a new collection of receptor conformers is formed based on the current. To form the new conformer collection, each validated ligand is docked into each conformer from current collection using standard VLS protocol (as implemented in ICM program), and for each ligand its best receptor conformers are selected (based on the ICM docking score). Then, the side chains of conformers in the vicinity of a bound ligand are optimized by biased probability Monte Carlo (BPMC) (see *Flexible Ligand Docking*). BPMC generates a stack of low energy most probable receptor/ligand complex conformations from which at least five are selected. These new conformations are appended to the current conformer collection. At this stage a new collection of receptor conformers can contain hundreds of members. Then, each ligand is docked using VLS protocol into each of the conformers, and new subensembles are formed for each ligand from its top scoring conformers. At least 25 receptor conformers were selected for each ligand. A new subensemble can contain both old and new receptor conformers; different subensembles can contain the same conformers. The conformers, which were not included into any of the subensembles, are eliminated.

Next, a new set of compounds for the next optimization round is compiled based on chemical fingerprint similarity (as implemented in the ICM program) to the current validated ligands. To increase the probability of the scaffold hopping event, the chemical search based on the chemical fingerprint is used instead of substructure-based search because the chemical fingerprint lacks the topological information. The ligand derivatives are searched in the complete ligand database (NCI database, ≈ 270,000 compounds). The top closest 100 ligand derivatives (based on a Tanimoto distance) are retained. The new compound set is then docked against updated receptor conformers using VLS protocol, and the top binders for each conformer

are retained (the score cutoff is applied). The hit lists for each conformer were then pooled together. At this point, the joint hit list might contain multiple entries of the same ligands if they scored high in multiple receptor conformers. The hit list was then sorted by the docking score and redundant ligand entries were eliminated. This procedure leaves us with the intersection of the top binders across updated collection of N1L conformers. The set of top predicted binders was further clustered by chemical similarity, and cluster representatives were ordered from NCI for *in vitro* screening.

The new *in vitro* validated ligand collection is formed as follows. If a derivative has better *in vitro* properties than the original validated ligand, then it is added to a new set of validated ligands. If none of the derivatives of a particular validated ligand has led to improvement of desired property, than this ligand, its derivatives and corresponding conformational subensemble are eliminated (Supplementary Figure 3B). The desired properties could be the IC_{50} values, cellular permeability and non-specific toxicity, protein tertiary structure stabilization by ligands (obtained by DSC), specificity (from cross-reactivity and DSC experiments) and so on. When a subensemble is deleted, its conformers are also deleted but only if they do not belong to any other existing subensemble. Then the procedure is repeated iteratively. The iterative application of the outlined above protocol results in both the structural refinement of putative induced-fit-like receptor conformations and the optimization of ligand scaffolds toward the desired *in vitro* properties.

Flexible Ligand Docking. In the ICM implementation, flexible ligand docking is performed via global optimization of the energy function [2,3]. The energy terms are based on the all-atom vacuum force field ECEPP/3 [4-6] with extra terms to account for solvation free energy and the

entropic contribution [7]. Modified intermolecular energy terms such as van der Waals and hydrogen bonding as well as a hydrophobic term may also be added [1].

The conformational search is based on the stochastic global optimization (SGO) procedure [8,9] which randomly selects a conformation in the internal coordinate space [10] and then makes a step to a new random position independent of the previous one, but according to a predefined continuous probability distribution. After each random step, a full local minimization is performed to improve the efficiency of the procedure. Since some energy terms have no analytical derivatives, a double-energy SGO method [10] circumvents this obstacle by minimizing the energy with respect to the differentiable terms but calculating the full energy with the non-differentiable terms. This double-energy scheme allows for the incorporation of complex energy terms, such as surface-based solvation energy into the global optimization process.

During the search for the energy minimum in the conformation space of the ligand-receptor system, the internal coordinates of the ligand as well as its position are allowed to change, and also side chains of the active site of the receptor are free to move. Each step of the algorithm consists of a random conformational change of one of two types, torsional and positional, followed by local minimization as described above. A torsional move involves randomization within a subspace of torsional angles. A positional move involves a pseudo-Brownian random translation and rotation of the ligand as whole [10]. The associated energy includes a van der Waals term, a hydrophobic term based on the solvent accessible surface buried upon binding, a solvation electrostatic term computed as a boundary-element solution of the Poisson equation, a hydrogen-bond terms and the entropic term [1].

Virtual Ligand Screening. The ICM VLS method was used as previously described [11]. The portion of the single N1L monomer within 5.0 A of the docked Bim peptide was selected for

ligand docking. Five grid potential maps, representing receptor selection, were generated. These maps accounted for the hydrophobic, heavy atom and hydrogen van der Waals interactions, hydrogen-bonding interactions and electrostatic potential. Ligand molecules were prepared for docking by energy minimization in the absence of the receptor, and the lowest energy conformations were used as starting points for simulations of docking to N1L potential maps by the ICM method [7,12]. The quality of ligand pose prediction was evaluated by assigning the docking score, generated by the ICM scoring function [13]. Since the ICM docking method has a stochastic element, the docking simulations were conducted at least three times to ensure convergence. The ligand conformations with the lowest (best) scores were retained. The docking score of –32.0 was considered statistically significant score cutoff.

Computational Facilities. The Scripps Research Institute maintains a large 64-bit Linux cluster, Garibaldi, on the La Jolla campus to support scientific computation and data processing. The Garibaldi cluster has a total of 3936 CPUs available for computations. Between local and shared disks the Garibaldi cluster has over 60 terabytes of disk space available for computational data. The Garibaldi cluster uses the Portable Batch System (PBS) for job queuing to ensure maximum system throughput.

References

1. Abagyan, R. & Totrov, M. High-throughput docking for lead generation. *Curr Opin Chem Biol* **5**, 375-382 (2001).
2. Abagyan, R. & Argos, P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *Journal of Molecular Biology* **225**, 519-532 (1992).
3. Abagyan, R.A. & Mazur, A.K. New methodology for computer-aided modelling of biomolecular structure and dynamics. 2. Local deformations and cycles. *Journal of Biomolecular Structure & Dynamics* **6**, 833-845 (1989).

4. Momany, F., McGuire, R., Yan, J. & Scheraga, H. Energy parameters in polypeptides. IV. Semiempirical molecular orbital calculations of conformational dependence of energy and partial charge in di- and tripeptides. *J Phys Chem* **75**, 2286-2297 (1975).
5. Nemethy, G., Pottle, M. & Scheraga, H. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occuring amino acids. *J Phys Chem* **87**, 1883-1887 (1983).
6. Nemethy, G. et al. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J Phys Chem* **96**, 6472-6484 (1992).
7. Totrov, M. & Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **Suppl 1**, 215-220 (1997).
8. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM: a new method for structure modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry* **15**, 488-506 (1994).
9. Abagyan, R. & Totrov, M. Ab initio folding of peptides by the optimal-bias Monte Carlo minimization procedure. *J Comp Chem* **151**, 402-421 (1999).
10. Abagyan, R. & Totrov, M. Biased Probability Monte-Carlo Conformational Searches and Electrostatic Calculations For Peptides and Proteins. *Journal of Molecular Biology* **235**, 983-1002 (1994).
11. Schapira, M., Abagyan, R. & Totrov, M. Nuclear hormone receptor targeted virtual screening. *Journal of Medicinal Chemistry* **46**, 3045-3059 (2003).
12. Totrov, M. & Abagyan, R. Protein-ligand docking as an energy optimization problem. in *Drug-receptor thermodynamics: Introduction and applications* (ed. Raffa, R.B.) 603-624 (John Wiley & Sons, Ltd., 2001).
13. Totrov, M., Abagyan, R. Derivation of Sensitive Discrimination Potential for Virtual Ligand Screening. in *Proceedings of the Third Annual International Conference on Computational Molecular Biology* (ed. Istrail, S., Pevzner, P., Waterman, M.) (Association for Computing Machinery, New York, 1999).