census included only virtual screens to identify small molecules interacting with protein targets. Protein target classes were not categorized for the census, nor was any differentiation made between virtual screens against protein crystal structures or homology models. The collection of reports was further filtered to include only prospective virtual screens to identify small molecules active against a particular protein target; validation studies were included only if they included some prospective lead-identification component. In many examples included in the census, additional techniques were used to prefilter compound databases or postscore docking results – indeed, it was only a rare occurrence when the screen practitioner did not at least postscore the results by visual inspection. Reports in which docking was the primary part of the virtual screening process were retained, but reports in which docking was only a minor adjunct to a virtual screen using pharmacophores, molecular shapes, substructures, or 2D descriptors were excluded. The actual docking programs used were not categorized, and the census represents every major and many minor docking programs.

Virtual screen results were extracted from the collection of reports that survived this filtering process. A surprisingly large number of peer-reviewed manuscripts reported no experimental data and instead reported only computed docking scores without any experimental verification of the proposed inhibition; all such reports were removed from the final tally. For the virtual screens with reported experimental data, scientists performing the docking studies selected a widely varying number of compounds for experimental assays, from as few as two molecules to more than 500. The final tally did not include every experimental result for every selected molecule from any particular virtual screen but instead included only the single best result for any molecule chosen in the virtual screening process. In some studies, hits were followed up by substructure searching or synthesis of related molecules; only the original hits from the virtual screen were included in the tally. In many cases, authors reported experimental results from the primary assay but had not performed orthogonal assays to confirm those results. In these instances I made no attempt to assess the plausibility of any particular result; no reported hits were excluded, even in cases where hit molecules contained features conducive to aggregation or assay interference. The literature search and filtering process described here of course could not produce a complete and exhaustive set of all manuscripts describing a prospective docking-based virtual screen, but the process has generated a sufficiently representative set of publications to allow for the analysis of trends in the real-life use of docking for lead identification.

Analysis of the census results

Table 7.1. Docking-based virtual screens for which experimental results were reported during the period January 2000–October 2008, aggregated into activity bins

Activity range	Count	Reference
<1 μM	18	41–58
1–10 μM	32	59–89
10–100 μM	37	90–126
>100 μM	9	104; 127–134
No hits	3	135–137

At the end of this literature search and filtering process, the final census included ninety-eight publications representing ninety-nine docking-based virtual screens for which experimental data were reported.[41–137] The best reported activity for each screen has been tabulated and binned into five activity categories: <1 μM, 1–10 μM, 10–100 μM, >100 μM, and no hits; these aggregate affinity results are listed in Table 7.1. The largest number of hits fell in the 10–100 μM range; for this category, the average IC_{50} value was 33 μM and the median value 24 μM. Eighteen virtual screening hits were reported to have experimental activities of <1 μM, which corresponds to two potent hits for each year in the census period. The census compilation contained very few results with affinities >100 μM, presumably due to reporting bias and lowered publication acceptance rates for negative results.

These aggregate data have been broken out by year (Figure 7.1); in this graph, the final two categories, ">100 μM" and "no hits," were combined into a single bin. As expected based on simple averaging, roughly two submicromolar hits per year were reported during the census period, with only two years early in the decade for which no such hits were reported. In 2007, the number of published docking-based virtual screens increased substantially, relative both to the previous year and to the pattern of increases over the period 2000–2005. Albeit with small changes in the actual distribution between the two low-IC_{50} categories, the number

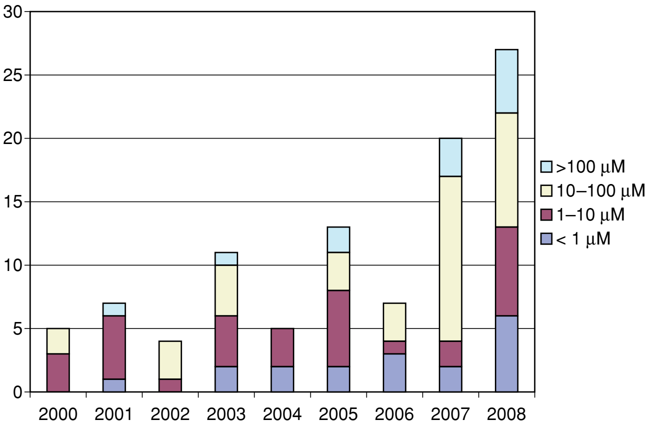

Figure 7.1. Docking-based virtual screens for which experimental results were reported during the period January 2000–July 2008, categorized by year.