J Comput Aided Mol Des (2010) 24:935–942
DOI 10.1007/s10822-010-9385-x

pK_a based protonation states and microspecies for protein–ligand docking

Tim ten Brink · Thomas E. Exner

Received: 24 June 2010 / Accepted: 16 September 2010 / Published online: 30 September 2010

Abstract In this paper we present our reworked approach to generate ligand protonation states with our structure preparation tool SPORES (Structure PrOtonation and REcognition System). SPORES can be used for the preprocessing of proteins and protein–ligand complexes as e.g. taken from the Protein Data Bank as well as for the setup of 3D ligand databases. It automatically assigns atom and bond types, generates different protonation, tautomeric states as well as different stereoisomers. In the revised version, pKa calculations with the ChemAxon software MARVIN are used either to determine the likeliness of a combinatorial generated protonation state or to determine the titrable atoms used in the combinatorial approach. Additionally, the MARVIN software is used to predict microspecies distributions of ligand molecules. Docking studies were performed with our recently introduced program PLANTS (Protein–Ligand ANT System) on all protomers resulting from the three different selection methods for the well established CCDC/ASTEX clean data set demonstrating the usefulness of especially the latter approach.

Keywords Molecular docking · Input generation · Protomers · ChemAxon MARVIN

Introduction

In recent years the use of protein–ligand docking has become a standard method in drug development. Despite this, a generally accepted protocol, when to use which docking tool but also how to setup the complex for docking, which gives good results on all kinds of protein targets or even non-protein targets, is still unavailable [1]. Some reports comparing docking tools showed that each of these tools has its strengths and weaknesses and that in general the approximation made in the scoring functions can be blamed for the success rates not exceeding much more than 80% and for the complete failure on some targets or even classes of targets [2–5]. But even if the same docking tool is used, different groups can come to different results on the same target. This is caused by different parameter settings in the docking runs but, even more important, by different setup procedures of the ligand and protein structures. E. g. Schrödinger Inc. gives a guarantee for reasonable docking results of their program GLIDE only if their setup procedure is exactly followed [6–9]. The effect of ligand conformation was systematically studied by Feher et al. [10] showing a considerable dependence on small structural changes. Additionally, since hydrogen atom positions are normally not available from X-ray structures of the complexes, methods for placing these atoms and for deciding on the protonation state of a ligand (and the protein) in its bound state have to be applied or this information has to be provided by the user. Again taking GLIDE as an example, the most probable protonation and tautomeric state is calculated with the LIGPREP module. But, the protein groups surrounding the ligand can highly influence the local pH, so that a different protonation could be favored in the bound state. To account for this effect, the ideal case would be to use multiple protonations in the docking and have the algorithm automatically pick the correct state. Recently this journal dedicated its issue from June 2010 Perspectives in Drug Discovery: Tautomers and Tautomerism to the problem of tautomers and protonation.

T. ten Brink · T. E. Exner (✉)
Department of Chemistry and Zukunftskolleg, University of Konstanz, 78457 Konstanz, Germany
e-mail: thomas.exner@uni-konstanz.de

In this issue the problems of tautomerism in general [11, 12], the problem of creating tautomer structures [13–16] and their use in drug design [17–19], especially in QSAR [14, 20, 21] and virtual screening [15] as well as the problems of tautomerism in large chemical databases [22] are noted. Another review about tautomers in drug design was given earlier by Martin [23]. They all come to the conclusion that tautomerism has been neglected in computational drug design for too long. Several software tools to generate ligand tautomers and protomers have been published. The Chemical Computing Group presented PROTONATE 3D [24] which adds hydrogen atoms and generates tautomers, rotamers and determines an accurate protonation for polar groups in the binding site using a free energy optimization. QUACPAC [25] from OpenEye Scientific Software is designed to give correct protonation states based on pK_a, tautomer enumeration and partial charge calculations. In contrast to QUACPAC, TAUTOMER, a tool by Molecular Networks, uses a rule based method for over 20 different forms of tautomerism. In the molecular modeling software package Sybyl [26] from Tripos the ProtoPlex plug-in can be used to generate protomers and tautomers of ligand molecules. An overview of software packages dealing with various problems regarding tautomers and their representation was given by Warr [27].

Despite the mentioned publications about protonation and tautomers in drug design, remarkably few publications deal with the problem of protomeric and tautomeric states in docking. Pospisil et al. [28] gave a short review about tautomers in drug design in which they concluded, that the inclusion of tautomers can be beneficial for virtual screening as the coverage of chemical space is improved and false negative hits are avoided. Brenk et al. [29] conducted a comparison of virtual screening and high throughput screening on the *E. coli* dihydrofolate reductase (DHFR) target, in which they emphasize the need of tautomeric states and protomers to avoid false negative predictions by overlooking likely protonation states or to create false positive predictions by including unrealistic protonation states. As recently shown, the enumeration of all possible protonations and the use of all these in the docking is contra-productive. Very often unreasonably highly-charged structures with wrong poses are generated nevertheless obtaining high scores due to the functional forms of today's scoring functions [30, 31]. A short time ago, we introduced our structure preparation tool SPORES [31] along with a study on the influence of the introduction of ligand protonation states into the docking process. A so-called standard protonation was defined based on a set of rules designed to give, at least in most cases, the protomer corresponding to physiological conditions. With this standard protonation, good docking results were obtained but it could not fully compete with the manually derived ligands. Using the enumerated protonation states resulted in a decrease in the success rate, which could be forced to the standard protonation results by adding a penalty term [31]. This can, as already mentioned, be attributed to unreasonably highly-charged protomers. The influence of different ligand tautomeric and protonation states on virtual screening was investigated by Kalliokoski et al. [30]. They also concluded that tautomer and protomer prediction can significantly save computing resources and can yield similar results to enumeration. Concluding, these studies demonstrate that unreasonable states should be removed prior to docking or must be additionally penalized. This can be done by calculating pK_a values and add an energy penalty for the protomers which are not the most favorable at physiological conditions as done in GLIDE[6, 8, 9] or in an approximative way as in the earlier version of SPORES [31].

But as just mentioned, we were not able to obtain superior results by using multiple protomers and penalizing unfavorable states compared to the standard protonation or even the manually generated protonations in our previous publication [31]. Here, we present an improved approach implemented in the new SPORES version. Different sophisticated filtering systems, utilizing the pK_a calculation capacity of the ChemAxon software MARVIN [32], are established to eliminate unlikely ligand protonations. It will be shown that in this way improved reproduction of the ASTEX clean data set [33] can be obtained with respect of pose prediction as well as identification of the correct protomer when using the PLANTS docking program [34, 35].

Materials and methods

Data set and structure preparation

The new results are reported for the same test set as used in our previous publication [31]: the ASTEX clean data set [33]. This set consists of 235 protein–ligand complexes, which are all of medical interest. A large effort has been invested to prepare the data set with all structures manually checked [33], so that we consider these structures to have the most favorable or in other words the "correct" protonation state in the bound form. Since the PLANTS software is not specially designed to cope with covalently bound complexes, all these were removed from the original data set resulting in the 213 complexes of the ASTEX $clean^{nc}$ data set. Since we are only dealing with the influences of the ligand preparation here, the protein structures were taken directly from the data set. For the ligand structures the automated structure preparation with our tool SPORES [31] was used. pK_a calculations and predictions

of microspecies distributions were done with the ChemAxon software MARVIN [32]. Since the atom and bond type convention of SPORES and the original data set cannot directly be used in MARVIN, a special input structure was generated for each ligand molecule. Three different docking studies were performed in which different methods of ligand protonation state generation and selection were tested.

Filtering of SPORES protonation states with pK_a values

In this first study the protonation states were generated with the original SPORES combinatorial method [31]. It is based on predefined functional groups and adds and removes single hydrogen atoms from these groups thus generating all combinatorial possible protonation states for the ligand (at least according to the defined groups). Protomers for the docking were then selected by pK_a values calculated with the CXCALC command line tool [32] of MARVIN. The groups were classified as very acidic, acidic, neutral, basic and very basic atoms based on specific pK_a thresholds. Protonation states were rejected if very acidic atoms were protonated or very basic atoms were deprotonated. A third criterion for rejection was the presence of a protonated acidic atom along with a deprotonated basic atom. Different thresholds were tested resulting in different protonation state distribution for the ligands as described later in the "Results and discussion".

Protonation states by marvin pK_a values

In the second study the pK_a calculation with CXCALC was used to determine the atoms suitable for a change in protonation instead of relying on the predefined groups of the previous approach. Each ligand atom was checked whether the pK_a value was below a certain acidic threshold or above a certain basic one. Atoms with pK_a values below the acidic threshold and above the basic threshold were treated as always deprotonated and protonated, respectively. On the other hand, for atoms in between these thresholds the combinatorial protomer generation method was applied. In this way, also groups not considered in the first approach can be changed if they fall in the appropriate pK_a range. Again the thresholds were varied to measure their influence on the docking results.

Microspecies distribution

The third study relies not on pK_a calculation but on the CXCALC prediction for the microspecies distribution of the ligands. As described in the introduction, the reason for using multiple protonation states is that the functional groups of the protein surrounding the ligand influence which protomer is the most stable one inside the binding site. This influence of the protein can be modeled by a shift in the local pH in the binding site. Negatively charged active sites would have a smaller local pH and positively charged active sites would have a higher local pH. But since the optimal pH for a specific site is very hard to estimate, the microspecies distribution was calculated at physiological $pH = 7.4$. After transferring all generated protomers to mol2 format, they were filtered so that only structures with more than 1% probability were retained. This guarantees that only reasonable structures are retained but gives the docking some flexibility to chose the best protomer according to the binding site conditions. This approach is reasonable (as shown below) for most proteins in the ASTEX data set, since the change of the local pH is only minor so that the optimal protomer is also included in the microspecies contribution at $pH = 7.4$. Only for metalloproteins a special treatment seems to be advantageous due to the large local charge concentration on the metal ion. For these cases, the pH value for the microspecies distribution was raised to $pH = 8.4$. In the last step the atom and bond types were readjusted with SPORES to give comparable structures to the previous studies.

Docking

All ligand structures were docked 25 times into their receptor using PLANTS with the ChemPLP scoring function and standard speed settings (SPEED 1) [36]. The docking was considered as successful if the best ranked structure with no regard of its protonation had a heavy atom root mean square deviation (RMSD) below 2 Å to the crystal structure.

Results and discussion

Original docking results

For comparison, the docking results previously obtained on the data set are briefly outlined. In its original protonation (ASTEX protonation) 170 of the 213 complexes were docked correctly resulting in a success rate of 80%. When the automatically generated standard protonation of our structure preparation tool SPORES was taken, this success rate dropped to 160 (76%). With the solely combinatorial protonation states a further reduction to 156 (73%) was observed. The lower success rate of the SPORES standard protonation can be explained by the adaptation of the ASTEX protonation to the protein binding site. While SPORES does not take any information on the protein into account [31], the protein and ligand protonation was adapted to optimal hydrogen bond interactions between the two

complex partners during the original setup of the ASTEX clean data set [33]. The further reduction of the success rate when the combinatorial generated protonation states are used can be explained by highly-charged states, which obtain very good scores because of their ability to form many charged interactions with the protein but have wrong poses.

Filtered SPORES protonation states

The success rates for the first study, where the combinatorial generated protonation states were filtered according the calculated pK_a values, depend, not surprisingly, on the threshold values used to classify the atoms. Three filter settings were tested (see Table 1).

The filter thresholds can be characterized by the amount of protonation states retained for the docking with the low and high setting removing the least and largest number of protomers, respectively. The high filter setting reached the best success rate. For the low setting, an overall success rate of 151 (71%) was obtained. This success rate rose to 153 (72%) and 155 (73%) when the medium and high settings are applied, respectively. Interestingly, the number of complexes, in which the ASTEX protonation is identified as the most favorable one, also increases in this order (see Table 2).

The filtering of the combinatorial generated protonation states resulted in a great modification of the distribution of protonation states per ligand. While the number of ligands with 16 or more different protonation states was 11 before the filtering, it dropped to 3 even with the lowest filter setting (Fig. 1a). With the highest filter setting, the maximum number of protonations states per ligand was 4. While this saves computation time, the lower overall success rate compared to the full enumerated set of protonation states (155 for the high setting compared to 156 correctly docked complexes for the full enumeration) is a drawback, even if this is probably not statistically significant. In some cases like the human carbonic anhydrase I (pdb code 1azm and 2h4n) failure can be explained by the exclusion of the correctly docked protonation state by the filter routine even at low filter settings. For other complexes (e. g. factor XA in complex with RPR208815, pdb code 1f0r) the reason seems to be that different protonation states gain very similar scores but have different poses. Due to the heuristic approach of PLANTS, not always the same minima are found and thus the ranking can be different. This sampling problem, even if we tried to minimize it by multiple docking runs, is not only limited to ligands with different protonation states but can also occur with larger ligands which have only a single protonation state (e.g. HIV protease in complex with VX-478, pdb code 1hpv). While the problems with unfavorable protonation states decrease when higher filter settings are used, another problem occurs when to many protonation states are excluded. Complexes, which contain metal ions in the binding site, often require ligand protonations adapted to the binding site. With the high filter settings these protonation states are excluded, which prevent a further increase in the overall success rate. Concluding this part, the fundamental problem of the combinatorial protonation states could only be solved to some extent using this filtering approach. The predefined functional groups limit the success of the method because it excludes some possible protonation states a priori.

A positive finding with the filtered protonation states is that with increasing filter setting the amount of top-ranked protonation states, which correspond to the original protonation in the data set, increases (see Table 2). This shows that the filtering results, at least partly, in a more chemically reasonable protonation of the bound ligand.

Table 1 The three filter intensities with the different pK_a thresholds used for atom classification

	Very acidic	Acidic	Basic	Very basic
Low	2	4	9	12
Medium	3	5	8	11
High	4	5	8	10

The filter setting "low" keeps most of the protonation sates while the filter setting "high" leads to the greatest reduction of protonation sates. The number gives the pK_a value used as threshold for the given atom class and filter setting

Table 2 The number of correctly docked complexes for the different filter settings

	Low	Medium	High
Correct	151	153	155
ASTEX	111	116	122
SPORES std	102	104	112

The number of correctly docked complexes with the best ranked ligand protonation corresponding to the original data set protonation is given as ASTEX and corresponding to the SPORES standard protonation is given as SPORES std

Protonation states based on MARVIN pK_a values

Here four different thresholds for variable atoms were tested. An overview of the thresholds is given in Table 3 and the results are summarized in Table 4. The lowest, in which all atoms with pK_a values between 2 and 12 were considered as variable, resulted in a success rate of 158 (74%) correctly docked complexes. When the thresholds were set to 3 and 11 the success rate decreased to 156 (73%). With a setting of 4 and 10, 158 (74%) complexes were correctly docked. A further restriction of the number of variable atoms (thresholds of 5 and 9) leads to a small

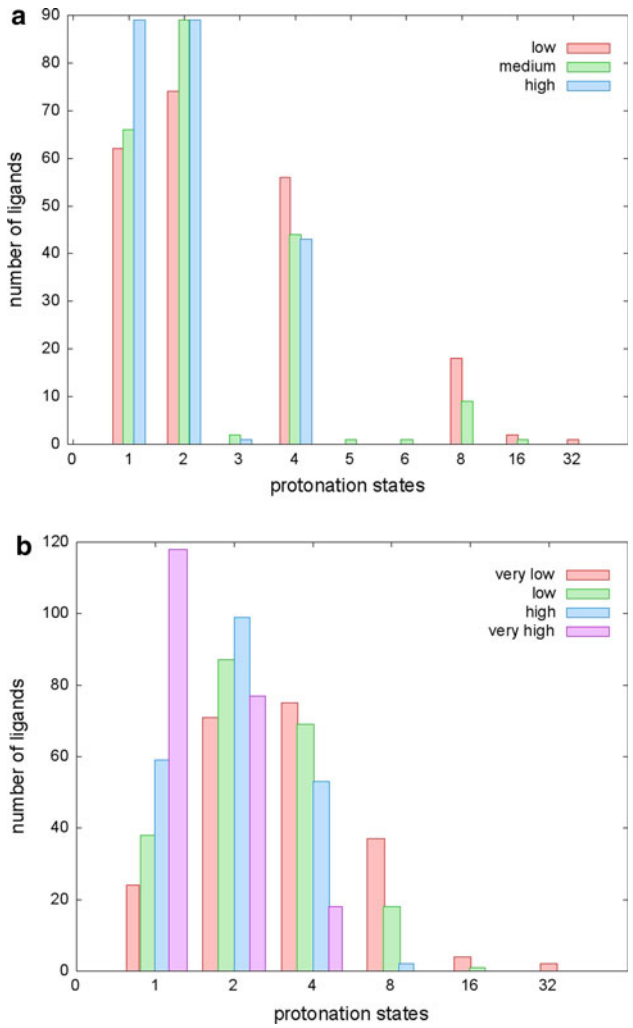

Fig. 1 Comparison of the protonation state distributions. **a** State distribution with the three different filter settings used on the original combinatorial protonation states. **b** State distribution for the four filter settings used for the protonation states directly based on the pK_a values

Table 3 The four filter intensities of the acidic pK_a limit below which an atom is considered as always protonated and the basic pKa limit above which an atom is considered always deprotonated

	Acidic limit	Basic limit
Very low	2	12
Low	3	11
High	4	10
Very high	5	9

Atoms with pKa values between these limits are considered as variable in their protonation

increase in the number of correctly docked complexes to 159 (75%). Some sampling problems seem to be the most obvious reason for the decrease in success rate from the lowest to the second lowest filter setting.

Overall, the different filters show very similar performances. But these are reached in much shorter computation times with the second highest and the highest filter setting due to the smaller number of states to be docked. The amount of protonation states per ligand strongly depends on the filter settings. A comparison of the distribution of the protonation states is given in Fig. 1b. With the filter settings very low and low more than 80% of the ligands have more than one protonation state. For the very high setting, no ligand has more than 4 protomers. Please note that, since other atoms as those of the predefined groups in the previous study can contribute to the different protonation states, they are not directly comparable to the distributions shown in Fig. 1a.

Microspecies distribution

Physiological value of pH = 7.4 was chosen despite the fact that some of the crystal structures of the ASTEX clean data set were solved at different experimental conditions. Here we are trying to describe the protein–ligand complex in its physiological pH environment and not in the artificial conditions used for optimal crystal growth. The original ASTEX data set was also created without taking the actual pH of the crystallization buffer into account. The main criterion for the original ASTEX protonation was an optimal hydrogen bonding pattern between the protein and the ligand [33].

The success rate obtained with the protonation states derived from the predicted microspecies distribution was 161 (76%), which is the second best success rate only outperformed by the manually generated ASTEX protonation. In our previous publication [31] adding a penalty term to the score of protomers not corresponding to the standard protonation had a positive effect. Thus, we tried if this is also the case for the microspecies distribution. Here the most populated species was taken as the reference. For all other protonation states a positive value is added to the scoring function if a hydrogen atom is missing or an additional hydrogen atom is present (in the ChemPLP negative values represent good solutions). The value added for each hydrogen atom was varied and Table 5 gives an overview of the obtained success rates. The best one of 164 (77%) correctly docked complexes could be obtained with a penalty of 1 scoring unit per missing or additional hydrogen atom.

Table 4 The number of correctly docked complexes for the four filter settings along with the number of best ranked correctly docked ligands in ASTEX and SPORES standard protonation

	Very low	Low	High	Very high
Correct	158	156	158	159
ASTEX	97	94	109	123
SPORES std	96	94	107	121

Table 5 The number of correctly docked complexes for the microspecies docking in dependence of the penalty term

	No penalty	1 × penalty	2 × penalty	3 × penalty	4 × penalty	5 × penalty	10 × penalty
Correct	161	164	163	163	163	163	161
Most populated	129	145	149	151	155	155	161
SPORES std	122	134	139	139	143	144	148
ASTEX	131	141	144	144	144	144	140

Also given is the number of correctly docked, best-ranked structures which correspond to the most populated form from the microspecies prediction, to the original ASTEX protonation and to the standard protonation of SPORES

These results are very promising since the overall success rate significantly exceeds the one obtained with the combinatorial protonation states and even the one obtained with the SPORES standard protonation. One remaining problem is that the original data set protonation, which is manually adapted to the protein binding site, is in 14 cases not among the generated protonation states. 6 of these 14 cases would require a more acidic pH value in the microspecies prediction while the other 8 would require a more basic value to re-include the ASTEX protonation in the distribution. The latter 8 cases are all metalloproteins in which the metal ions drastically change the local pH in the binding site. A further investigation for all 213 complexes of the ASTEX data set showed that in 36 cases metal ions have direct contacts with the ligands in the crystal structure. In all but the mentioned 8 cases, the ligand protonations needed to correctly coordinated the ions are included in the micro species distribution predicted for pH = 7.4. An examination of the pK_a values of the metal coordinating atoms of the problematic ligands showed that in 7 of these 8 cases the ASTEX protonation is among the predicted protonation states when pH = 8.4 is used for the microspecies prediction (even if it only contributes with just above 1%). Only the protonation of the phenylamidazole ligand complexed to cytochrome P450 (pdb code 1phd) would require a pH value of nearly 12 to be included in the predicted protonation states. New docking runs using the microspecies distribution at pH = 8.4 show no change in the success rate but the number of ASTEX protonation is increased by 4 cases when no penalty term is applied. Fig. 2 shows the ligand and binding site of blLAP (pdb code 1lcp) as an example for a metal containing binding site in which the correct protonation is only included in the predicted microspecies distribution at pH = 8.4 or larger.

Finally, we would like to note that the automated generated SPORES standard protonation compares remarkably well to the manual revised ASTEX protonation in this study. In only 3 cases the SPORES standard protonation is not among the predicted protonation states. Although the ASTEX protonation corresponds more often to the best ranked structure as shown in Table 5, the SPORES standard protonation is in 192 cases identical to the most populated structure in the microspecies distribution and the ASTEX protonation only in 177 cases. This shows that our standard protonation is a very good approximation to the most favorable structure at physiological conditions.

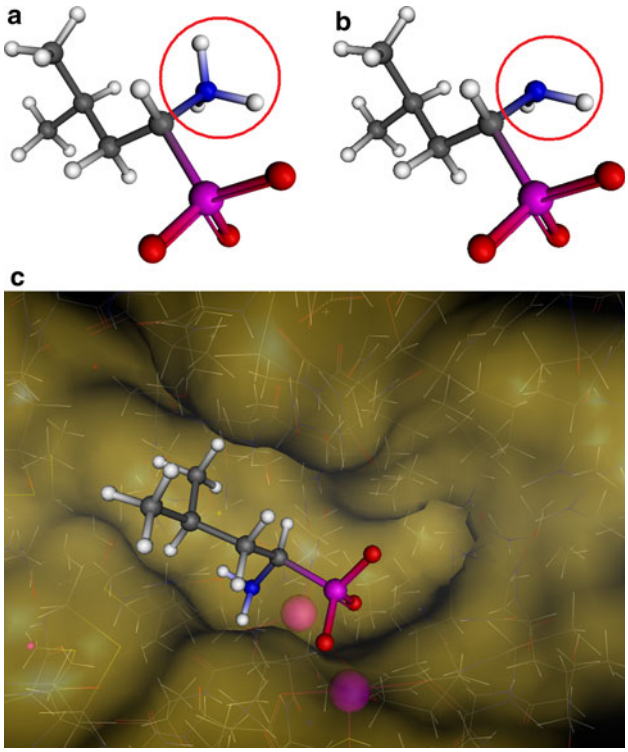

Fig. 2 **a** The SPORES standard protonation of the leucine phosphonic acid ligand of pdb entry 1lcp: This protomer corresponds to the most populated microspecies MARVIN predicts at pH = 7.4. **b** The original ASTEX protonation of the same ligand: It is only included in the microspecies distribution if pH > = 8.4 is used for the prediction. **c** The binding mode of the ligand (ball and stick model) in the crystal structure where the ligand is coordinating two zinc ions (cpk model) in the binding site (wire frame model with surface), one with the phosphate and one with the amino group. This indicates that the ASTEX protonation is correct for the bound state of the ligand

Conclusion and outlook

As summary, all success rates of our previous study [31] and the ones of the best performing filter setting for each approach described here are listed in Table 6. The similar

Table 6 Summarization of the docking success rates

	Success rate all	Success rate non-metalloproteins	Success rate metalloproteins	Number of dockings
ASTEX	170 (80%)	80%	77%	213
SPORES std	160 (76%)	77%	64%	213
Combinatorial	156 (73%)	73%	72%	938
pK_a filtered	155 (73%)	73%	72%	406
pK_a based	159 (75%)	76%	67%	344
Microspecies	164 (77%)	78%	72%	367

The success rates are listed separately for the whole data set, for the subset of the 177 non-metalloproteins and for the 36 metalloproteins. The number of dockings is the total number of ligand protomers contained in the study. For each method of protonation state generation only the best setting in terms of success rate is listed

performance regarding success rates of the combinatorial generated protonation states and the ones generated on the basis of pK_a stands in good agree with the study of Kalliokoski et al. [30], who compared the performance of enumerated to predicted protonation and tautomeric states on virtual screening experiments. Nevertheless, our new approaches have a number of advantages rendering them useful as preprocessing step for docking. First and most important, docking with the protonation states generation based on the prediction of microspecies distributions outperforms the other methods including the purely combinatorial approach from the previous study. Only the correct protonation directly taken from the ASTEX data set gives better results. But the manual adaptation of the protonation state to the conditions in the active site, as done in the latter procedure, is only possible if the complex structure is already known. Second, the filtering largely decreases the number of protomers which have to be docked resulting in a considerable time saving (the number of needed docking runs for each approach is also given in Table 6). The number of docking runs is only increased by 50% using the microspecies distribution while more than 4 times the number of docking runs is needed for the combinatorial approach. Third, the problem with highly-charged, unstable protomers scored highest can be removed. And last but not least, the use of penalties for microspecies with lower probability can increase the success rate even further and at the same time guides the docking to identify the correct protomers in the bound form. Due to the uncertainties in the pK_a calculations, we decided not to use a penalty dependent on the pK_a values as done in [6, 8, 9]. Instead just a constant value for adding or deleting a proton compared to the most populated protomer is applied.

An interesting finding is the performance of the most populated protonation state of the ligands predicted for $pH = 7.4$. Even without the penalty term which presses the ranking toward this structure, 80% of the correctly docked ligands and 78% of all ligands have the best ranked structure in this protonation state. This shows that for the majority of the proteins the ligand protonation has not to be adapted to the binding site and explains the good performance of the standard protonation. For automated ligand preparation tools, which do not take the binding site into account like SPORES, this leads to the conclusion that a good representation of the ligand protonation at physiological pH can result in docking success rates comparable to manually revised structures at least for standard protein targets. Only metal-ion-containing binding sites have been identified here to be in need of a special treatment since the ions have a great influence on the ligand protonation. Thus, it is beneficial to include some information on the binding site e.g. by modifying the pH value for which the microspecies distribution is calculated. Overall, the metalloproteins show a reduced docking success rate compared to the non-metalloproteins from the data set in all approaches (see Table 6, the subset of the metalloproteins includes the 36 complexes, where the ligand is interacting with the ion in the crystal structure). This is especially the case for the standard protonation as well as high filter settings (approach 1: pK_a based), in which the correct protomer is very often not among the ones used for docking. But if the protonation can adopt more flexibly to the binding site (ASTEX and approach 3: microspecies), results between metalloproteins and non-metalloproteins are much more comparable

Acknowledgments We would like to thank ChemAxon for the MARVIN software provided within the ChemAxon's free Academic Package program and Oliver Korb for helpful discussions and support for the PLANTS software. We also thank the Common Ulm Stuttgart Server (CUSS) and the Baden-Württemberg grid (bwGRiD), which is part of the D-Grid system, for providing the computer resources making the simulations possible.

References

1. Moitessier N, Englebienne P, Lee D, Lawandi J, CC R (2008) Towards the development of universial, fast and highly accurate docking/scoring methods: a long way to go. BJP 153:7–26. doi: 10.1038/sj.bjp.0707515

2. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 57:225–242
3. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 56:235–249
4. Kontoyianni M, McClellan LM, Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. J Med Chem 47:558–565
5. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) Ligandfit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph 21:289–307
6. Glide (2005) 4.0 edn. Schrödinger, LLC, New York, NY
7. Maestro (2006) 7.5 edn. Schrödinger, LLC, New York, NY
8. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1 method and assessment of docking accuracy. J Med Chem 47(7):1739–1749. doi:10.1021/jm0306430
9. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: A new approach for rapid, accurate docking and scoring. 2 enrichment factors in database screening. J Med Chem 47(7):1750–1759. doi:10.1021/jm030644s
10. Feher M, Williams CI (2009) Effect of input differences on the results of docking calculations. J Chem Inf Model 49(7): 1704–1714. doi:10.1021/ci9000629
11. Sayle R (2010) So you think you understand tautomerism? J Comp Aided Mol Des 24(6):485–496
12. Klamt A, Diedenhofen M (2010) Some conclusions regarding the predictions of tautomeric equilibria in solution based on the sampl2 challenge. J Comp Aided Mol Des 24(6):621–625
13. Clark T (2010) Tautomers and reference 3d-structures: The orphans of in silico drug design. J Comp Aided Mol Des 24(6):605–611
14. Cramer R (2010) Tautomers and topomers: challenging the uncertainties of direct physicochemical modeling. J Comp Aided Mol Des 24(6):617–620
15. Greenwood J, Calkins D, Sullivan A, Shelley J (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comp Aided Mol Des 24(6):591–604
16. Haranczyk M, Gutowski M (2010) Combinatorial–computational–chemoinformatics (c3) approach to finding and analyzing low-energy tautomers. J Comp Aided Mol Des 24(6):627–638
17. Henry R (2010) The effects of tautomerism on the nature of molecules in the solid state. J Comp Aided Mol Des 24(6): 587–590
18. Katritzky A, Hall C, El-Gendy B, Draghici B (2010) Tautomerism in drug discovery. J Comp Aided Mol Des 24(6):475–484
19. Porter W (2010) Warfarin: history, tautomerism and activity. J Comp Aided Mol Des 24(6):553–573
20. Cruz-Cabeza A, Schreyer A, Pitt W (2010) Annular tautomerism: experimental observations and quantum mechanics calculations. J Comp Aided Mol Des 24(6):575–586
21. Martin Y (2010) Tautomerism, hammett σ, and qsar. J Comp Aided Mol Des 24(6):613–616
22. Sitzmann M, Ihlenfeldt W-D, Nicklaus M (2010) Tautomerism in large databases. J Comp Aided Mol Des 24(6):521–551
23. Martin Y (2009) Let's not forget tautomers. J Comp Aided Mol Des 23(10):693–704
24. Labute P (2009) Protonate3d: Assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins 75:187–205
25. Quacpac (2008) 1.3.1 edn. OpenEye Scientific Software, Inc
26. Sybyl (2008) 7.2 edn. Tripos Inc., 1699 South Hanley Road, St. Louis, Missouri, 63144, USA
27. Warr W (2010) Tautomerism in chemical information management systems. J Comp Aided Mol Des 24(6):497–520
28. Pospisil P, Ballmer P, Scapozza L, Folkers G (2003) Tautomerism in computer-aided drug design. J Recept Signal Transduct 23(4):361–371. doi:10.1081/RRS-120026975
29. Brenk R, Irwin JJ, Shoichet BK (2005) Here be dragons: docking and screening in an uncharted region of chemical space. J Biomol Screen 10(7):667–674. doi:10.1177/1087057105281047
30. Kalliokoski T, Salo HS, Lahtela-Kakkonen M, Poso A (2009) The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. J Chem Inf Model 49(12): 2742–2748. doi:10.1021/ci900364w
31. ten Brink T, Exner TE (2009) Influence of protonation, tautomeric, and stereoisomeric states on protein–ligand docking results. J Chem Inf Model 49(6):1535–1546. doi:10.1021/ci800420z
32. Marvin (2009) 5.2.0 edn. ChemAxon
33. Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R (2002) A new test set for validating predictions of protein–ligand interaction. Proteins 49:457–471
34. Korb O, Stützle T, Exner TE (2006) Plants: application of ant colony optimization to structure-based drug design. In: Dorigo M, Gambardella LM, Birattari M, Martinoli A, Poli R, Stützle T (Eds) Ant Colony Optimization and Swarm Intelligence, 5th international workshop, ANTS 2006, LNCS 4150:247–258
35. Korb O, Stützle T, Exner TE (2007) An ant colony optimization approach to flexible protein-ligand docking. Swarm Intell 2(1):115–134
36. Korb O, Stützle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with plants. J Chem Inf Model 49(1):84–96. doi:10.1021/ci800298z10.1021/ci800298z