

Contents

| | |
|-----------------------------------------------------------|-------------|
| List of Tables | IV |
| List of Figures | V |
| List of Abbreviations and Symbols | VIII |
| 1 Summary | 1 |
| 2 Zusammenfassung | 3 |
| 3 Introduction | 5 |
| 3.1 Enzymes | 6 |
| 3.2 Activity modulation using small molecules | 7 |
| 3.3 Drug development | 10 |
| 3.4 Drug design | 12 |
| 3.5 Computer assisted drug discovery | 14 |
| 3.5.1 Data complexity and <i>big data</i> | 15 |
| 3.5.2 Virtual screening | 17 |
| 3.5.3 Pocket identification | 18 |
| 3.5.4 Molecular surface | 20 |
| 3.6 Fractals | 21 |
| 3.6.1 Analytical complexity | 24 |
| 3.6.2 Scale invariance | 25 |
| 3.6.3 Self-affinity and self-similarity | 27 |
| 3.6.4 Fractals in nature | 32 |
| 3.6.5 Fractal dimension | 33 |
| 3.7 Fractal analysis of macromolecular surfaces | 38 |
| 3.8 Project goals | 40 |

| | | |
|----------|------------------------------------------------------------------------|-----------|
| 4 | Materials and methods | 42 |
| 4.1 | Software base | 42 |
| 4.1.1 | C++ | 42 |
| 4.1.2 | Solvent accessible surface calculation with MSMS | 43 |
| 4.1.3 | Statistical modelling with R | 44 |
| 4.1.4 | Machine learning software | 44 |
| 4.1.5 | NAMD | 47 |
| 4.1.6 | MODELLER | 48 |
| 4.2 | Molecular databases | 49 |
| 4.2.1 | A protein-ligand complex database: PDBind | 50 |
| 4.2.2 | A refined protein-ligand complex dataset: the <i>ligandability</i> set | 50 |
| 4.2.3 | The sc-PDB protein-ligand complex database | 51 |
| 4.3 | Targets | 51 |
| 4.3.1 | 4-Diphosphocytidyl-2-C-methyl-D-erythritol synthase | 53 |
| 4.3.2 | uPA | 56 |
| 4.3.3 | Glucokinase | 58 |
| 4.3.4 | Interleukin-2 | 69 |
| 4.3.5 | HIV-1 protease | 70 |
| 5 | Results | 74 |
| 5.1 | Roughness of molecular surfaces | 74 |
| 5.2 | <i>LoRI</i> : Local Roughness Indicator | 80 |
| 5.3 | Ligandability prediction of pharmaceutical targets | 85 |
| 5.3.1 | uPA | 86 |
| 5.3.2 | IspD | 86 |
| 5.3.3 | Interleukin-2 | 88 |
| 5.3.4 | Glucokinase | 89 |
| 5.3.5 | HIV protease | 91 |

| | |
|----------------------------------------------------------------------|------------|
| 6 Discussion | 95 |
| 6.1 Parametrization of the roughness calculation | 95 |
| 6.2 Roughness of biologically active molecules and patches | 96 |
| 6.3 <i>LoRI</i> | 97 |
| 6.4 Ligandability prediction of pharmaceutical targets | 98 |
| 6.4.1 uPA | 98 |
| 6.4.2 IspD | 99 |
| 6.4.3 IL-2 | 99 |
| 6.4.4 Glucokinase | 99 |
| 6.4.5 HIV protease | 101 |
| 6.5 General considerations | 102 |
| 7 Conclusion and outlook | 104 |
| 8 Acknowledgements | 106 |
| Appendices | 126 |
| A Tables | 126 |
| B Curriculum Vitae | 130 |