Supporting Information

Protein Electrostatic Properties Predefine the Level of Surface Hydrophobicity Change upon Phosphorylation

Anton A. Polyansky†§, Bojan Zagrovic*†§

†Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna, AT-1030, Austria, and ‡M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russia, and §Mediterranean Institute for Life Sciences, Split, Croatia, and †Department of Physics Faculty of Science, University of Split, Split, Croatia

E-mail: bojan.zagrovic@univie.ac.at
### Table S1. Modeled protein pairs.

| Protein                      | Abbr. | UniProt       | Source                        | PDB id     | Method  | Fragment | N res | Mod. Res |
|------------------------------|-------|---------------|-------------------------------|------------|---------|----------|-------|----------|
| **Globular**                 |       |               |                               |            |         |          |       |          |
| Phosphocarrier protein Hpr   | HPR   | P07515        | Enterococcus faecalis         | 1FU0 / 1PTF| X-ray   | whole    | 87    | Ser46    |
| Endo-1,4-beta-xylanase Y     | Xyn10B| P51584        | Clostridium thermocellum      | 1GKK / 1GKL| X-ray   | 792-1077 | 283   | Ser954   |
| Putative anti-sigma factor antagonist TM1442 | TM1442 | Q9X1F5 | Thermotoga maritima | 1T6R / 1SBO | NMR | whole | 110 | Ser59    |
| HP-like protein Crh          | CRH   | Q06976        | Bacillus subtilis             | 2AK7 / 1MU4| X-ray   | whole    | 86    | Ser46    |
| Alkaline phosphatase, placental type | PLAP  | P05187        | Homo sapiens                  | 2GLQ / 1ZED| X-ray   | 23-506   | 479   | Ser92    |
| STE20-like serine/threonine-protein kinase | hSLK  | Q9H2G2        | Homo sapiens                  | 2JFL / 2J51| X-ray   | 19-320   | 288   | Thr183 / Ser189 |
| **Disordered**               |       |               |                               |            |         |          |       |          |
| Oxoglutarate dehydrogenase inhibitor | Odh   | QBNQ13        | Corynebacterium glutamicum    | 2XB3 / 2XB4| NMR    | whole    | 143   | Thr15    |
| Protein phosphatase 1 regulatory subunit 14A | CPI17 | O18734        | Sus scrofa                    | 2RTL1 / 1J2M| NMR   | 22-120   | 99    | Thr17    |

### Table S2. Hydrophobic properties of proteins during MD simulations.

| Protein                      | MHP\(^{\text{asa}}\)\(_{\text{ph}}\) | MHP\(^{\text{asa}}\)\(_{\text{nat}}\) | ΔMHP\(^{\text{asa}}\) |
|------------------------------|----------------|----------------|----------------|
| HPR                          | -246.5±112.0 | -243.1±62.2 | -3.3±92.3 |
| Xyn10B                       | -492.9±81.3  | -222.7±116.7| -270.2±134.4|
| TM1442                       | -767.8±65.9  | -579.5±74.2 | -188.3±100.6|
| CRH                          | -407.2±42.3  | -394.8±46.1 | -12.3±65.9 |
| PLAP                         | -2039.2±128.9| -1748.5±153.2| -290.7±183.2|
| hSLK                         | -782.8±100.8 | -746.6±132.1| -36.2±170.3|
| OdhI                         | -883.5±62.6  | -1043.2±87.8| 159.7±115.9|
| CPI17                        | -1290.2±83.3 | -736.7±89.8 | -553.5±133.7|

* MHP\(^{\text{asa}}\) – sum of MHP values on protein surface (distributions of MHP\(^{\text{asa}}\) from simulations in native and phosphorylated forms for all proteins, except HPR (p=0.13), are significantly different with p-values obtained according Mann-Whitney test < 2.2x10\(^{-16}\); ΔMHP\(^{\text{asa}}\) = MHP\(^{\text{asa}}\)\(_{\text{ph}}\) - MHP\(^{\text{asa}}\)\(_{\text{nat}}\).
Table S3. Settings of MD simulations.

| Protein | System* phospho | System native | Box size, Å³ |
|---------|-----------------|---------------|-------------|
| HPR     | 1/2566/6        | 1/2569/4      | 45x45x45    |
| Xyn10B  | 1/5703/8        | 1/5697/6      | 60x60x60    |
| TM1442  | 1/2424/3        | 1/2415/1      | 45x45x45    |
| CRH     | 1/6705/5        | 1/6686/3      | 60x60x60    |
| PLAP    | 1/9031/10       | 1/9046/8      | 70x70x70    |
| hSLK    | 1/9840/19       | 1/9851/15     | 70x70x70    |
| OdhI    | 1/16393/10      | 1/16370/8     | 80x80x80    |
| CPI17   | 1/2518/3        | 1/2480/1      | 45x45x45    |

* System – number of PROTEINS/SPC water molecules/Na⁺ counterions

**Extraction of the studied structure set.** All structures were extracted from the RCSB Protein Data Bank using an Advanced Search procedure. Final structures were selected using the following requirements: 1) the phosphorylated residues are explicitly present in the 3D structure; 2) both forms have the same length and sequence; 3) the molecules represent a complete protein or its biologically relevant fragment (e.g. isolated catalytic domains), 4) they do not contain missing parts (in the case of X-ray structure), 5) they are not membrane proteins, and 6) are not in complex with any binding partners.

**MD simulation details.** MD simulations were performed using Gromacs 4.0.7 package. For all systems (Table S3), the standard protocol was used. Proteins were placed in water boxes, together with the necessary amount of sodium counterions to reach neutrality, and subjected to energy minimization, followed by heating to 300 K for 100 ps and 150 ns of unconstrained MD simulations. All proteins were modeled using the Gromos 96 (43a1P) force field, including the parameters of phosphorylated residues. This force field is available via GROMACS website (www.gromacs.org), and has been developed on the basis of Gromos 96 (43a1) using partial-charges and van-der-Waals parameters of phosphorylated residues published elsewhere. It was also successfully applied in MD simulations of phosphorylated peptides. The SPC model was used for water molecules. MD simulations were carried out with a time step of 2 fs, with imposed 3D periodic boundary conditions, in the isothermal-isobaric (NPT) ensemble with an isotropic pressure of 1 bar and a constant temperature of 300 K. The pressure and the temperature were scaled using the Berendsen thermo- and barostat with 1.0 and 0.1 ps relaxation parameters, respectively. The van der Waals and electrostatic interactions were truncated using the twin range 10 / 12 Å spherical cutoff.

**Analysis of hydrophobic/hydrophilic properties of protein surface.** Hydrophobic properties of protein solvent accessible surface (SAS) were analyzed using the
molecular hydrophobicity potential (MHP) approach. The formalism of MHP is based on empirical atomic hydrophobicity constants (i.e., “hydrophobicity weights”) derived form partition coefficients, Log $P$, of various compounds between polar and apolar media (e.g. water / n-octanol). Although the MHP approach is partially limited by its empirical nature, the lack of hydrophobicity constants for a number of atom types and ions, and ambiguous assignment of hydrophobicity values for some particular compounds, it provides a unique tool for the prediction of LogP values for small-weight compounds, analysis of spatial hydrophobic properties of membrane and globular proteins, and quantitative assessment of the efficiency of protein-ligand interactions. In analogy with the electrostatic Coulomb potential, MHP is constructed to have distance dependence, which is typically exponential. Thus, contribution of $N$ atoms to MHP at point $i$ can be estimated as follows:

$$ M_{HP_i} = \sum_{j}^N f_j \times \exp(-c \times R_{ij}), $$

where $f_j$ is atomic hydrophobicity constant of atom $j$, $R_{ij}$ is the distance between atom $j$ and point $i$, and $c$ is a decay constant (here we used $c$ of 0.5 Å$^6$). SAS calculation and mapping of MHP onto protein surfaces in each of its points were performed using PLATINUM software. Further analysis of MHP data was carried out using utilities written especially for this. The MHP values were expressed in octanol/water Log $P$ values (base-10 logarithm of octanol/water partition coefficients). The sum of MHP values on protein surface (MHP$sas$) was used as a quantity of protein hydrophobicity. To estimate the difference in this value between phosphorylated and native states:

$$ \Delta M_{HP}sas = M_{HP_{ph}}sas - M_{HP_{nat}}sas, $$

where $M_{HP_{ph}}sas$ and $M_{HP_{nat}}sas$ refer to phosphorylated and native proteins, respectively. These values were calculated over the last 135 ns of MD with time separation of 100 ps. For analysis of local MHP properties at a phosphorylation site, all SAS points within 6.5 Å away from any atom in the modified residue were taking into account.

**Analysis of electrostatic potential on the protein surface.** Grid values of electrostatic potential ($V_E$) of MD protein snapshots were calculated over the last 135 ns of MD with time separation of 100 ps using APBS software. For this purpose, a modified AMBER 99 force-field including parameters for phosphorylated residues was employed. All calculations were carried out at 300 K, using 2 and 78 solute and solvent dielectric constants, and 0.55 Å grid spacing. The obtained grids were used to get values of $V_E$ at the same SAS points ($V_{E,sas}$) as MHP. Phosphorylation-induced changes of electrostatic potential on the surface were estimated as:

$$ \Delta V_{E,sas} = V_{E(ph),sas} - V_{E(nat),sas} $$
For the analysis of local electrostatic properties at a phosphorylation site, all SAS points belonging to a 6.5 Å – radius sphere from the modified residue were taking into account.

Visualization of MHP and $V_e$ was performed using PyMOL (http://www.pymol.org/).

**Analysis of contacts between charged residues.**

The all-against-all distance matrices for all charged residues in a given protein were obtained over the last 135 ns of MD with time separation of 1 ps using $g_{saltbr}$ utility from the GROMACS package. The numbers of negative-negative, positive-positive, and positive-negative pairs were calculated for all residues separated by less than 6.5 Å. These values were used to estimate the ratio of the each type of contacts in native and phosphorylated states of a protein.
**Figure S1.** Correlation of the relative MHP scale for amino acid sidechains with a hydrophobicity scale based on the partition $\Delta G$ values in the octanol-water mixture. The calculated values of MHP correspond to sums of atomic hydrophobicity constants (in octanol-water Log P units) for amino-acid sidechains. A value for GLY represents hydrophobicity of the backbone. The values of the octanol-water partition $\Delta G$ values for pSER, pTHR and pTYR are predictions based on the calculated MHP values and the depicted regression line.

**Figure S2.** Backbone RMSD of MD conformations from initial structures. Dashed lines depict a start point for the analysis of MD data.
Figure S3. (A) Representative MD conformations of native and phosphorylated forms of PLAP protein. The protein is shown with solvent accessible surface (SAS), which is colored according to values of electrostatic potential (Vₑ) in volts. The phospho-sites are shown with green dashed lines. (B) Phosphorylation-induced changes of electrostatic potential of SAS at the phospho-sites (open bars) and for the whole molecule (filled bars).
References

(1) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J. Chem. Theory Comput. 2008, 4, 435.

(2) Hansson, T. J. Mol. Biol. 1997, 265, 118.

(3) Bjarnadottir, U.; Nielsen, J. E. Biopolymers 2011.

(4) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration Reidel, Dordrecht, 1981.

(5) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J. Chem. Phys. 1984, 81, 3684.

(6) Efremov, R. G.; Chugunov, A. O.; Pyrkov, T. V.; Priestle, J. P.; Arseniev, A. S.; Jacoby, E. Curr Med Chem 2007, 14, 393.

(7) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. J. Phys. Chem. A 1998, 102, 3762.

(8) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. J. Comput. Aided. Mol. Des 1994, 8, 83.

(9) Pyrkov, T. V.; Chugunov, A. O.; Krylov, N. A.; Nolde, D. E.; Efremov, R. G. Bioinformatics 2009, 25, 1201.

(10) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Proc. Natl. Acad. Sci. U S A 2001, 98, 10037.

(11) Homeyer, N.; Horn, A. H.; Lanig, H.; Sticht, H. J. Mol. Model. 2006, 12, 281.

(12) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. Nature 2003, 424, 805.