J|A|C|S
ARTICLES
Published on Web 11/21/2006

Structure-Based Secondary Structure-Independent Approach To Design Protein Ligands: Application to the Design of Kv1.2 Potassium Channel Blockers

C. Magis, D. Gasparini, A. Lecoq, M. H. Le Du, E. Stura, J. B. Charbonnier, G. Mourier, J.-C. Boulain, L. Pardo, A. Caruana, A. Joly, M. Lefranc, M. Masella, A. Menez, and P. Cuniasse*

Contribution from the Département d'Ingénierie et d'Etude des Protéines, DSV, CEA, CE-Saclay, 91191 Gif Sur Yvette Cedex, France

Received June 30, 2006; E-mail: philippe.cuniasse@cea.fr

Abstract: We have developed a structure-based approach to the design of protein ligands. This approach is based on the transfer of a functional binding motif of amino acids, often referred as to the "hot spot", on a host protein able to reproduce the functional topology of these residues. The scaffolds were identified by a systematic in silico search in the Protein Data Bank for proteins possessing a group of residues in a topology similar to that adopted by the functional motif in a reference ligand of known 3D structure. In contrast to previously reported studies, this search is independent of the particular secondary structure supporting the functional motif. To take into account the global properties of the host protein, two additional criteria were taken into account in the selection process: (1) Only those scaffolds sterically compatible with the positioning of the functional motif as observed in a reference complex model were retained. (2) Host proteins displaying electrostatic potentials, in the region of the transferred functional motif, similar to that of the reference ligand were selected. This approach was applied to the development of protein ligands of the Kv1.2 channel using BgK, a small protein isolated from the sea anemone *Bunodosoma granulifera,* as the reference ligand. Four proteins obtained by this approach were produced for experimental evaluation. The X-ray structure of one of these proteins was determined to check for similarity of the transferred functional motif with the structure it adopts in the reference ligand. Three of these protein ligands bind the Kv1.2 channel with inhibition constants of 0.5, 1.5, and 1.6 μM. Several mutants of these designed protein ligands gave binding results consistent with the presumed binding mode. These results show that protein ligands can be designed by transferring a binding motif on a protein host selected to reproduce the functional topology of this motif, irrespective to the secondary structure supporting the functional motif, if the host protein possesses steric and electrostatic properties compatible with the binding to the target. This result opens the way to the design of protein ligands by taking advantage of the considerable structural repertoire of the Protein Data Bank.

Introduction

Increasing knowledge of the molecular aspects of biological processes paints a complex picture where a large number of identified receptors or enzymes interact with other proteins to achieve or control their functions. The development of modulators of these interactions is regarded as a formidable goal. Many approaches have been proposed to design such compounds, but the growing availability of three-dimensional (3D) structures of proteins offers many opportunities for the development of structure-based ligands, which is of paramount importance. The aim of these approaches is twofold: (1) to obtain compounds able to interact with receptors in order to discover novel therapeutics and (2) to tackle the fundamental basis of ligand−receptor interactions.

The concept of the "hot spot" of binding energy has recently emerged from studies aimed at understanding the principles of protein−protein interactions. The hot spot of binding energy is defined as the residues of a ligand that, when replaced by an alanine, give a significant reduction in the free energy of binding to the receptor considered.[1] It has been pointed out that the drop in binding constant observed when replacing a particular residue by an alanine should be interpreted with caution. Strictly speaking, this drop comes from the thermodynamic destabilization of the bound ensemble as compared to the unbound one and may come from many effects not always related to the suppression of direct interactions between the substituted residue and the receptor.[2] However, this technique has been extensively used to investigate the energetic of protein−protein interactions and, in a number of cases, allowed researchers to identify residues playing a predominant role in the binding energetic. The first demonstration of the existence of a hot spot of binding energy by alanine scanning was shown in the interaction between the human growth hormone and its receptor[1] and was

(1) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383−386.
(2) Delano, W. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14−20.

10.1021/ja0646491 CCC: $33.50

later generalized to many binding interfaces.[3,4] These works showed that the free energy was not evenly distributed at the protein−protein interface but that a limited number of residues of the ligand, the so-called hot spot, contributed most of the binding free energy. Analysis of clustered protein−protein interfaces demonstrated the conservation of some polar residues at specific locations.[5] A correlation was observed between structurally conserved residues at the protein−protein interface and experimental alanine scanning hot spots,[5,6] and computational hot spots.[7] In addition, it was shown that experimental hot spots and the conserved residue tend to couple across two-chain interfaces.[8] These works suggested that the structure of the interaction region is composed of a high-packing-density core surrounded by a low-packing-density rim of less conserved residues, the hot spot being located in the former. This view converged with the previously proposed O-ring model[3] as well as the results of other analyses of protein−protein recognition sites.[7,9,10] According to the O-ring models, the residues of the rim could play a role in solvent occlusion and therefore protect the interactions established by the hot-spot residues from destabilization. A similar effect can be expected from the high packing density of the core region as it certainly facilitates water exclusion upon binding. Overall, this mechanism is possible because the residues forming the hot spots are able to established specific interactions, such as hydrogen bonds or salt bridges with the protein partner.[7] The primary condition for these interactions to be established is that the residues forming the hot spot are in a defined topology. This led to the idea to design protein ligands by transferring a hot spot onto a scaffold able to reproduce the functional topology of the residues composing this motif.

Small proteins, sometimes referred to as mini-proteins,[11,12] have several characteristics that render them attractive for the development of ligands. Higher theoretical affinities can be expected, as the binding of such molecules will be accompanied by a smaller loss in entropy as compared to that observed with flexible peptides. Structural studies of mini-proteins in the free state can be easily undertaken by NMR due to their moderate size and can give meaningful data regarding their conformation in the bound state. Their small size renders mini-proteins accessible to chemical synthesis, and so it is possible to produce large amounts and to introduce unnatural amino acids. In addition, one may expect mini-proteins to be more stable than peptides with regard to the proteolytic activity they encounter in vivo. From a structural point of view, mini-proteins can be used as scaffolds to control the topology of the group of residues or chemical functions identified in a reference ligand as playing a predominant role in the interaction with the target. In most studies, the hot spot has been transferred onto a scaffold reproducing the same structural element, such as a beta hairpin, loop, or helix, as that supporting the functional motif in the reference ligand.[13−20] Many examples exist of hormones able to bind multiple receptors or, conversely, of protein receptors binding several hormones.[21] The Fc fragment of immunoglobulin G was shown to be able to bind at least four proteins, each having radically different folds but presenting a similar set of contact residues. In addition, it has been shown that a peptide selected in vitro to bind the Fc fragment of immunoglobulin G mimicked the interactions of the natural protein ligands but with a completely different structural fold as compared to the natural ligands.[22] On the basis of these observations, we reasoned that a protein ligand could be designed by transferring the group of functionally important residues onto scaffolds with secondary structure unrelated to that of the reference ligand but possessing a group of residues of topology similar to that of the functional motif in this reference ligand. Host proteins could be identified by a systematic search of the considerable structural repertoire offered by the Protein Data Bank (PDB).[23] Such an approach may have many advantages over the standard one, as it offers the opportunity to transfer the functional motif originally composed of residues present on discontinuous structural elements. In addition, the systematic examination of the structural repertoire of the PDB might lead to numerous structural solutions for the transfer of the considered hot spot, greatly improving the yield of the approach.

To be successful, the structure-based approach described above must take into account other important aspects of the protein−protein interaction. The global properties of the scaffolds have generally been disregarded in previous studies of protein ligand design. Many studies reporting the analysis of X-ray structures of protein−protein complexes improved our understanding of protein recognition in terms of global parameters and specific intermolecular interactions.[24−38] One very important aspect of protein recognition and interaction is the

(3) Bogan, A. A.; Thorn, K., S. *J. Mol. Biol.* **1998**, *280*, 1−9.
(4) Thorn, K. S.; Bogan, A. A. *Bioinformatics* **2001**, *17*, 284−285.
(5) Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2000**, *39*, 331−342.
(6) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5772−5777.
(7) Keskin, O.; Ma, B.; Nussinov, R. *J. Mol. Biol.* **2005**, *345*, 1281−1294.
(8) Halperin, I.; Wolfson, H.; Nussinov, R. *Structure* **2004**, *12*, 1027−1038.
(9) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177−2198.
(10) Chakrabarti, P.; Janin, J. *Proteins: Struct., Funct. Genet.* **2002**, *47*, 334−343.
(11) Cunnigham, B. C.; Wells, J. A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 457−462.
(12) Martin, L.; Vita, C. *Curr. Protein Pept. Sci.* **2000**, *1*, 403−430.
(13) Smith, M. B. K.; Rouzer, C. A.; Taneyhill, L. A.; Smith, N. A.; Hughes, S. H. Boyer, P. L.; Janssen, P. A. J.; Moereels, H.; Koymans, L.; Arnold, E.; Ding, J. P.; Das, K.; Zhang, W. Y.; Michejda, C. J.; Smith, R. H. *Protein Sci.* **1995**, *4*, 2203−2222.
(14) Drakopoulou, E.; ZinnJustin, S.; Guenneugues, M.; Gilquin, B.; Menez, A.; Vita, C. *J. Biol. Chem.* **1996**, *271*, 11979−11987.
(15) Domingues, H.; Cregut, D.; Sebald, W.; Oschkinat, H.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 652−656.
(16) Kellenberger, E.; Mer, G.; Kellenberger, C.; Marguerie, G.; Lefevre, J. F. *Eur. J. Biochem.* **1999**, *260*, 810−817.
(17) Vita, C.; Drakopoulou, E.; Vizzavona, J.; Rochette, S.; Martin, L.; Menez, A.; Roumestand, C.; Yang, Y. S.; Ylisastigui, L.; Benjouad, A.; Gluckman, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13091−13096.
(18) Chin, J. W.; Schepartz, A. *Andgew. Chem.* **2001**, *113*, 3922−3925.
(19) Golemi-Kotra, D.; Mahaffy, R.; Footer, M. J.; Holtzman, J. H.; Pollard, T. D.; Theriot, J. A.; Schepartz, A. *J. Am. Chem. Soc.* **2004**, *126*, 4−5.
(20) Gemperli, A. C.; Rutledge, S. E.; Maranda, A.; Schepartz, A. *J. Am. Chem. Soc.* **2005**, *127*, 1596−1597.
(21) Wells, J. A.;de Vos, A. M. *Annu. Rev. Biochem.* **1996**, 65, 609−634.
(22) Delano, W. L.; Ultsch, M. H.; de Vos, A. M.; Wells, J. A. *Science* **2000**, *287*, 1279−1283.
(23) Berman, H. M.; et al. *Acta. Crystallogr. D: Biol. Crystallogr.* **2002**, *58*, 899−907.
(24) Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705−708.
(25) Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C. *Nature* **1987**, *328*, 834−836.
(26) Argos, P. *Protein Eng.* **1988**, *2*, 101−113.
(27) Janin, J.; Chothia, C. *J. Biol. Chem.* **1990**, *265*, 16027−16030.
(28) Jones, S.; Thornton, J. M. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 31−65.
(29) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13−20.
(30) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *J. Mol. Biol.* **1996**, *260*, 604−620.
(31) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121−132.
(32) Larsen, T. A.; Olson, A. J.; Goodsell, D. S. *Structure* **1998**, *6*, 421−427.
(33) Jones, S.; Thornton, J. M. In *Protein−Protein Recognition*; Kleantheous, C., Ed.; Oxford University Press: Oxford, UK, 2000; pp 33−59.
(34) Ponstingl, H.; Henrick, K.; Thornton, J. M. *Proteins: Struct., Funct. Genet.* **2000**, *41*, 47−57.
(35) Glaser, F.; Steinberg, D. M.; Vakser, I. A.; Ben-Tal, N. *Proteins: Struct., Funct. Genet.* **2001**, *43*, 89−102.
(36) Ofran, Y.; Rost, B. *J. Mol. Biol.* **2003**, *325*, 377−387.

shape complementarity of the interacting partners, the most obvious aspect of which is the absence of steric hindrance. Several criteria have been proposed to characterize shape complementarity.[39,40] This aspect must be explicitly taken into account in the selection of protein hosts able to interact with the considered target. However, this is not the sole criterion in predicting the correct binding mode. Electrostatics plays a considerable role in protein−protein interactions[41−44] via either direct electrostatic interactions or solvation. The electrostatic potential of a protein is determined by all of its amino acids and depends on their topology. In the context of using proteins as scaffolds to transfer a functional motif, the particular electrostatic potential of the scaffold could give either a favorable or an unfavorable contribution to the binding and must therefore be taken into account explicitly.

The aim of the present work is to evaluate the possibility to design protein ligands by transferring a binding motif of amino acids on a protein host able to reproduce its functional topology as defined in a reference ligand. Specifically, the identification of these protein hosts is achieved by a systematic search through the Protein Data Bank, irrespective of the secondary structure of the ligand of reference. Global properties of these scaffolds are explicitly taken into account in the selection process. This includes screening to select proteins that do not exhibit any steric hindrance with the targeted receptor in a binding mode that reproduces the position of the binding motif in the structure of the reference ligand−receptor complex. The resulting proteins are then filtered to retain only those scaffolds that reproduce an electrostatic potential in the region of the binding motif similar to that of the reference ligand. This structure-based approach was applied to the design of protein ligands able to block the potassium channel Kv1.2.

Results and Discussion

The Interaction Studied. In order to evaluate our approach to structure-based design by hot-spot transfer, we chose as targets the Kv1 potassium channels, which can be blocked by natural toxins from sea anemones. Among these natural blockers, BgK, a 37-amino-acid protein isolated from the sea anemone *Bunodosoma granulifera*, has been extensively studied. Alanine scanning substitution of BgK showed that residues S23, K25, and Y26 constitute this toxin's functional core in its binding to the Kv1.1 and Kv1.2 channels.[45,46] K25 and Y26 constitute the functional dyad of BgK that is shared by many natural channel blockers from scorpions, sea anemones, and snakes.[47,48] Replacement of residue S23 by an alanine residue in BgK leads to a significant decrease in binding to both Kv1.1 and Kv1.2 channels, while the substitution of residue F6 by an alanine results in a significant reduction in binding to Kv1.2 and a moderate drop in binding to Kv1.1. We selected the motif composed of residues F6, S23, K25, and Y26 (hereafter called FSKY motif) as a hot spot to transfer onto different scaffolds, so as to design new compounds able to interact with the Kv1.2 channels.

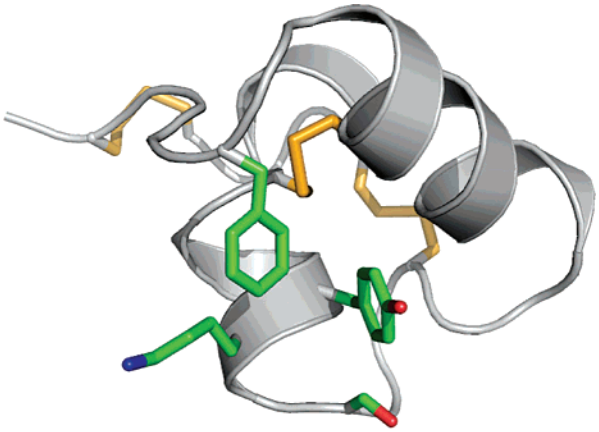


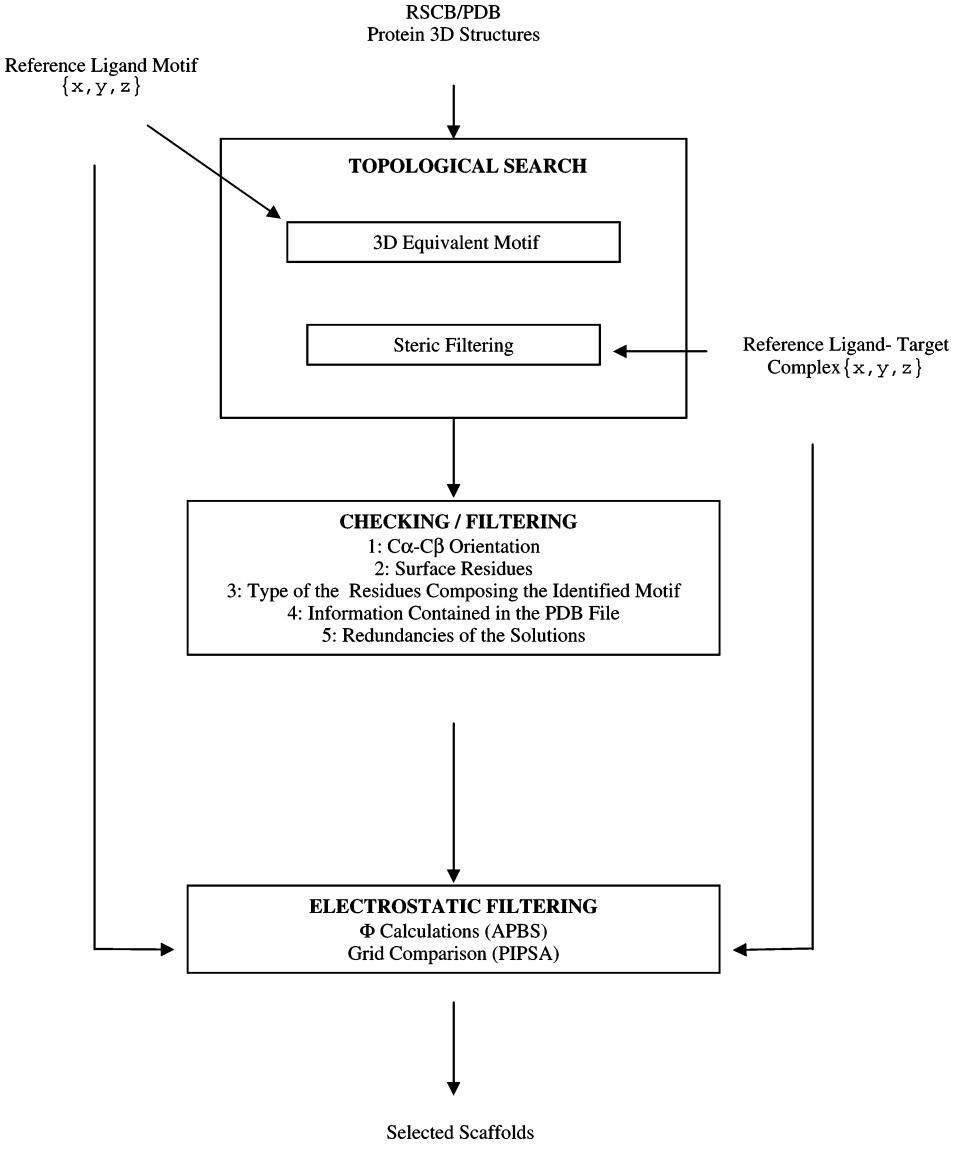
Figure 1. Schematic representation of the average minimized 3D structure of BgK used as reference in this study. The main chain is shown in cartoon representation. The residues composing the F6, S23, K25, and Y26 (FSKY) motif are shown in stick representation.

Structure of the Functional Motif. It has long been recognized that the topology adopted by a hot spot is critical to its functional behavior. The functional topology adopted by the FSKY motif can be defined via the 3D structure of BgK as determined by NMR.[47] Examination of the 3D structure of BgK shows that residues S23, K25, and Y26 are located on the N-terminal side of the second α helix observed in the structure, while residue F6 is located in the loop preceding the first α helix. While forming a compact motif, the residues of the hot spot are separated within the amino acid sequence of the protein and located in different structural elements. Therefore, this hot spot is a good test case because, in contrast to most previously reported studies, our method is based on the identification of protein scaffolds possessing a group of residues in a topology similar to that adopted by the hot spot in the reference structure, irrespective of the secondary structure of the scaffold. This can be achieved by a systematic examination of the protein structures deposited in the Protein Data Bank[23] with the topological constraints deduced from the structure of the reference ligand. In the present case, this search was used to identify proteins with a group of four residues adopting a topology similar to that observed for the FSKY motif in BgK (Figure 1).

An in Silico Stepwise Approach. Our protein ligand design approach can be described in three steps, summarized in Chart 1. The initial step consists of a systematic topological search for scaffolds in the PDB and includes a steric filtering of the solutions to take into account the structure of the target. During this step, an upper limit for the size of the proteins analyzed as scaffolds can be set. During the second step, several checking and filtering steps are achieved. Checking concerns the topology of the identified motif as compared to the reference one, the presence of the motif's residues at the surface of the scaffold, the structural information actually contained in the PDB files

(37) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. *J. Mol. Biol.* **2004**, *336*, 943−955.
(38) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. *Proteins: Struct., Funct. Genet.* **2003**, *53*, 708−719.
(39) Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, 234, 946−950.
(40) Laskowsky, R. A. *J. Mol. Graph.* **1995**, *13*, 323−330.
(41) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct. Genet.* **1988**, *4*, 7−18.
(42) Vakser, I. A.; Aflalo, C. *Proteins: Struct., Funct. Genet.* **1994**, *20*, 320−329.
(43) Jackson, R. M.; Sternberg, M. J. E. *J. Mol. Biol.* **1995**, *250*, 258−275.
(44) Weng, Z. P.; Vajda, S.; Delisi, C. *Protein Sci.* **1996**, *5*, 614−626.
(45) Alessandri-Haber, N.; Lecoq, A.; Gasparini, S.; Grangier-Macmath, G.; Jacquet, G.; Harvey, A. L.; de Medeiros, C.; Rowan, E. G.; Gola, M.; Menez, A.; Crest, M. *J. Biol. Chem.* **1999**, *274*, 35653−35661.
(46) Gilquin, B.; Racape, J.; Wrisch, A.; Visan, V.; Lecoq, A.; Grissmer, S.; Menez, A.; Gasparini, S. *J. Biol. Chem.* **2002**, *277*, 37406−37413.
(47) Dauplais, M.; Lecoq, A.; Song, J. X.; Cotton, J.; Jamin, N.; Gilquin, B.; Roumestand, C.; Vita, C.; deMedeiros, C. L. C.; Rowan, E. G.; Harvey, A. L.; Menez, A. *J. Biol. Chem.* **1997**, *272*, 4302−4309.
(48) Gasparini, S.; Danse, J. M.; Lecoq, A.; Pinkasfeld, S.; Zinn-Justin, S.; Young, L. C.; de Medeiros, C. C. L.; Rowan, E. G.; Harvey, A. L.; Menez, A. *J. Biol. Chem.* **1998**, *273*, 25393−25403.

Chart 1



of the selected structures, and the possible redundancies of the solutions. In this step, filtering concerns the minimal size of the protein to be considered as a scaffold and the type of native amino acids composing the identified motif. In a third step, after transfer of the hot-spot residues, these solutions are submitted to an electrostatic filtering step were the electrostatic potential of the scaffold is compared to that of the reference ligand.

Topological Search for Host Proteins in the PDB. As our objective is to transfer hot-spot residues on a protein scaffold, the search must be achieved without taking into account the particular type of residues composing the native motif in the scaffold but taking into account the orientation of the side chains. Therefore, we systematically searched the protein structures in the PDB for a motif of Cα,Cβ carbons topologically similar to those of the FSKY motif residues (Chart 1). The structure of the hot spot of BgK was therefore characterized by a set of interatomic distances separating all the Cα and Cβ carbons of the four residues in the motif, calculated from the average minimized NMR structure of BgK (Experimental Section; Table 1). The search was done with a program originally developed in our group for several applications, with the aim of identifying any set of atoms or chemical functions adopting a defined topology in the structures of proteins deposited in the protein database (unpublished results). This program has similarities with the SPASM program,[49] including the description of the 3D motif with a distance matrix and the recursive search through the atomic coordinates. We also implemented a maximum deviation criterion (Δd_MAX) between the interatomic distances

Table 1. Interatomic Distances between the Cα and Cβ of the Residues Composing the FSKY Motif in the Average Minimized Structure of BgK.

		F6		S23		K25		Y26	
		Cα	Cβ	Cα	Cβ	Cα	Cβ	Cα	Cβ
F6	Cα								
	Cβ	1.52							
S23	Cα	13.37	12.37						
	Cβ	12.41	11.31	1.52					
K25	Cα	8.84	8.31	5.87	5.76				
	Cβ	8.13	7.56	6.57	6.24	1.50			
Y26	Cα	8.81	7.90	5.04	4.30	3.78	4.45		
	Cβ	9.79	8.73	4.10	3.08	4.64	5.16	1.50	

(49) Kleywegt, G. J. *J. Mol. Biol.* **1999**, *285*, 1887−1897.

describing the search motif and the corresponding one found in a particular protein. Similarly to SPASM, if at least one of these distances is greater than the value set for Δd_MAX, the motif is rejected. The topological similarity is characterized by the root-mean-square deviation (rmsd) of the atomic coordinates in the identified motif as compared to those of the motif sought after superimposition of the two motifs. If the value of the rmsd is greater than the threshold value RMSD_MAX, the structure is rejected. However, our program has some specific features, some of them being essential for the application reported in the present paper. The first one is dictated by the search for scaffolds. As pointed out above, we do not search for topological motifs composed of similar residues as those of the motif to transfer. Thus, we searched for Cα,Cβ carbons in a topology similar to that adopted by the FSKY motif residues. This necessitates that we use a PDB version containing the coordinates of all the Cα and Cβ atoms of the proteins. SPASM uses a simplified representation of the structures (each residue of a protein in the PDB is represented by the coordinates of its Cα and a pseudo-atom corresponding to the center of gravity of its side chain). Our program works on a full atomic representation of the 3D structures of the proteins, therefore containing all Cα and Cβ coordinates of the deposited structures.

There is another very important option, central to the present work, that has been implemented in our program: the steric filter (Chart 1). When selecting a scaffold able to reproduce the topology of a particular motif like the FSKY motif of BgK, depending on the location of the motif on the scaffold and on the particular structure of this scaffold, other structural elements present in the protein may lead to some steric hindrance that might preclude the correct positioning of the motif relative to the target. When the structure of the complex between the reference ligand and the target is known, it is possible to filter the scaffolds on which to transfer the binding motif to avoid such steric hindrance. This steric filter has been implemented in our program. As input data, we used the 3D structure of the complex between the target and the reference ligand. Each scaffold identified on the basis of the presence of a Cα,Cβ motif topologically equivalent to the one sought is superimposed on the corresponding atoms of BgK in the complex with the target. The distances between all non-hydrogen atoms of the scaffold and all non-hydrogen atoms of the target are then calculated. This is possible because our program works on a full representation of the 3D structures of the PDB. In the absence of steric hindrance between the target and the scaffold, all distances must be greater than the sum of the radii of the corresponding atoms. Conversely, in the presence of steric hindrance, one may observe distances below the sum of these radii. As our search treats the structures of the scaffold and the target as rigid, there may be some steric hindrance involving the side chains in any region of the two proteins that might be easily relaxed by some minor reorientations of these side chains. To take this into account, we implemented a user-definable threshold value for the maximum percentage of atoms of the target that display at least one distance violation with an atom of the scaffold (%D_MAX). We counted the number of atoms of the target to allow for the comparison of scaffolds of different size. Therefore, %D_MAX gives the maximum percentage of non-hydrogen atoms of the target that can overlap the tested scaffold. When this percentage is below the threshold value, it is concluded that there is no significant steric hindrance.

During our search for scaffolds, we analyze all conformers deposited in the NMR structures. This option was chosen because there is substantial conformational variability in the NMR structures and we found that many solutions were discarded due to the threshold effect when we analyzed only the first conformer deposited, as is the case with SPASM (data not shown).

Parameters for the Topological Search. The systematic search for scaffolds able to reproduce the topology of the FSKY motif as observed in BgK was conducted in the full PDB[23] (2004 edition). Several constraints were applied for the screening of the database. Scaffolds possessing fewer than 20 or more than 70 residues were rejected. Small peptides (fewer than 20 residues) are generally flexible and can hardly be considered to behave as scaffolds. The exclusion of proteins possessing more than 70 residues is based on the fact that we search for mini-proteins accessible to peptide synthesis. The amino acid composition of the motif present in the identified scaffold was subject to some restrictions. Scaffolds possessing the native residues proline and cysteine were not selected. Indeed, these residues are well known to play a particular role in defining or stabilizing the 3D structure of the proteins.[50] Replacement of these residues by the corresponding one in the FSKY motif might have a large impact on the 3D structure adopted by the scaffold after substitution. For similar reasons, we limited the number of glycine residues to one in the amino acid composition of the native motif. Redundant solutions were eliminated. This included identical proteins found in several different PDB files or solutions found in several conformers reported in NMR-derived structures. In these cases, the structures selected were those that displayed the lowest rmsd from the Cα,Cβ atoms of the native motif as compared to the corresponding atoms in the FSKY motif sought.

BgK, the reference ligand in the present study, binds with similar high affinity to Kv1.1, Kv1.2, and Kv1.6 channels.[51] These channels share high sequence homology, particularly in the region of the S5−S6 part of the channel where BgK binds. A model of the interaction between the S5−S6 region of Kv1.1 with BgK has recently been reported.[46] This model was derived from a docking procedure based on back-calculated coupling free energy obtained from double mutant cycles. This work allowed well-defined positioning of the functional core (S23, K25, Y26) of BgK in Kv1.1. The main feature of the interaction of BgK with the Kv1.1 is the location of the K25 residue that enters the pore of the channel. This feature has also been recently reported in the NMR-derived structure of a related Kcsa modified channel with charybdotoxin that possesses a similar functional core with a lysine residue entering the pore of the channel.[52] Thus, the BgK/S5−S6 modeled structure was used in the present study as reference complex for the steric hindrance filtering of the scaffold identified by the systematic analysis of the protein structure in the PDB.

The search for scaffolds able to reproduce the topology of the FSKY functional core was achieved with an RMSD_MAX

(50) Chou, P. Y.; Fasman, G. D. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1978**, *47*, 45−148.
(51) Racape, J.; Lecoq, A.; Romi-Lebrun, R.; Liu, J.; Kohler, M.; Garcia, M. L.; Menez, A.; Gasparini, S. *J. Biol. Chem.* **2002**, *277*, 3886−3893.
(52) Yu, L.; Sun, C.; Song, D.; Shen, J.; Xu, N.; Gunesekara, A.; Hajduk, P. J.; Olejniak, E. T. *Biochemistry* **2005**, *44*, 15834−15841.

Table 2. Scaffolds Possessing a Cα,Cβ Motif Corresponding to the FSKY Motif of BgK, Identified by a Systematic Search through the PDB[a]

PDB code	name	RMSD	FSKY equivalent residue				chain ID
			F	S	K	Y	
1BGK	BgK[47]	0.12	6	23	25	26	
1BNB	bovine neutrophil β-defensin[53]	0.83	32	1	3	4	
1CKK	calmodulin-dependent protein kinase[54]	0.64	26	6	8	9	B
1D6B	defensin-like peptide 2[55]	0.62	20	34	36	37	A
1EDL	staphylococcal protein AE domain[56]	0.69	39	4	6	7	
1H5O	crotamine[57]	0.63	15	31	33	34	A
1UXD	fructose repressor DNA-binding domain[58]	0.83	28	13	15	16	
1J75	DLM1[59]	0.35	132	162	164	165	A
1OQE	s TALL-1/BAFF-R[60]	0.64	21	5	7	8	O
1BOE	IGF-binding domain of IGFBP-5[61]	0.57	57	28	30	31	A
1HZ6	B1 domain of protein L[62]	0.74	26	53	55	56	C
1J9E	NT2[63]	0.65	30	46	48	49	A
1A7F	insulin mutant[64]	0.92	8	1	4	5	A
2PDD	domain of dihydrolipoamide acetyltransferase[65]	0.77	1	25	27	29	
2SH1	neurotoxin 1 from sea anemone[66]	0.77	20	29	46	30	
1ERF	Nter fusion peptide of HIV-1 GP4[67]	0.70	22	2	4	5	A
1F7E	EGF-like domain[68]	0.64	87	73	75	76	A
1M3A	Nter SH3 domain c-Crk[69]	0.53	169	147	149	150	A
1EGF	murine epidermal growth gactor[70]	0.67	43	10	12	13	
1EAY	CHEY-binding domain of CHEA[71]	0.99	215	170	174	171	C
1FGP	protein G3P[72]	0.78	21	47	49	50	

[a] RMSD is the root-mean-square deviation from the position of the Cα,Cβ atoms after superimposition of the identified motif in the scaffold on the corresponding atoms in the FSKY motif of BgK (in Å). The chain identity and the positions of the residues topologically equivalent to the FSKY motif are indicated.

value set to 1.0 Å and a Δd_MAX value of 1.3 Å. These values were chosen after several runs of the program and examination of the resulting solutions. Values of RMSD_MAX below 1.0 Å resulted in very few solutions identified in the full PDB. In addition, one should notice that using values of RMSD_MAX below 1.0 Å would implicitly overestimate the precision of the structures derived by NMR or even by X-ray diffraction. The same argument applies to the value of Δd_MAX. In addition, we observed that reducing the value of Δd_MAX significantly below 1.3 Å tended to return very few solutions. Increasing Δd_MAX above 1.5 Å led to a significant fraction of solutions that displayed some unacceptable distortions of the resulting topology as compared to the one sought.

Identification of Protein Scaffold for the FSKY Motif of BgK. The search for scaffolds able to reproduce the topology of the FSKY motif of BgK in the full PDB (23 845 structures) with the parameters described above returned 8291 solutions corresponding to proteins possessing fewer than 70 residues. This large number is explained by the fact that, at this step, the search can return any motif of four residues composed of any type of residue with no more than one glycine. The location of the motifs found can be either on the surface or in the core of the protein. Furthermore, the size of the motif is small (four residues), and the constraints concern only the Cα and Cβ atoms. It should be noted that the number of solutions returned by the topological search depends on the number of residues taken into account in the hot spot. Increasing the number of residues increases the number of constraints and therefore should decrease the number of solutions. However, preliminary tests of the methods on other systems suggested that the number of solutions returned depends not only on the number of residues but also strongly on the particular topology adopted by the residues taken into account (data not shown).

In the present application, selection of motifs present at the surface of the scaffolds can be achieved with steric hindrance filtering. In the case of a motif located in the core of the scaffold, its correct positioning in the reference complex must lead to steric overlap between the scaffold and the Kv1 channel. The parameter %D_MAX (see above) used for the steric hindrance filter was set to 2%. This value was empirically calibrated via several runs of the program and examination of the solutions. When %D_MAX values greater than 2% were used, in some cases the resulting scaffolds displayed unacceptable steric overlap involving their main chain with that of the Kv1.1 channel. It must be noted that the current implementation of our steric filter may have one drawback. Indeed, we cannot exclude that some solutions rejected because they correspond to more than 2% of the atoms of the target overlapping with the considered scaffold actually involve only side chains and therefore could be satisfactory solutions. However, our main goal in the present work is to describe the approach and to reach a diversity of solutions that can be tested to evaluate the potential of the method.

In the present case, application of the steric filter led to rejection of about 90% of the solutions obtained in the absence of steric filtering that were not sterically compatible for the interaction with the Kv1 channel when the scaffolds were positioned by superimposing the motif identified on the FSKY motif in BgK (887 solutions selected). Exclusion of the redundant structures of scaffolds possessing either a proline or a cysteine residue in any position of the motif in the identified

(53) Zimmermann, G. R.; Legault, P.; Selsted, M. E.; Pardi, A. *Biochemistry* **1995**, *34*, 13663−13671.
(54) Osawa, M.; Tokumitsu, H.; Swindells, M. B.; Kurihara, H.; Orita, M.; Shibanuma, T.; Furuya, T.; Ikura, M. *Nat. Struct. Biol.* **1999**, *6*, 819−824.
(55) Torres, A. M.; de Plater, G. M.; Doverskog, M.; Birinyi-Strachan, L. C.; Nicholson, G. M.; Gallagher, C. H.; Kuchel, P. W. *Biochem. J.* **2000**, *348*, 649−656.
(56) Starovasnik, M. A.; Skelton, N. J.; O'Connell, M. P.; Kelley, R. F.; Reilly, D.; Fairbrother, W. J. *Biochemistry* **1996**, *35*, 15558−15569.
(57) Nicastro, G.; Franzoni, L.; De Chiara, C.; Mancin, A. C.; Giglio, J. R.; Spisni, A. *Eur. J. Biochem.* **2003**, *270*, 1969−1979.
(58) Penin, F.; Geourjon, C.; Montserret, R.; Bockmann, A.; Lesage, A.; Yang, Y. S.; Bonod-Bidaud, C.; Cortay, J. C.; Negre, D.; Cozzone, A. J.; Deleage, G. *J. Mol. Biol.* **1997**, *270*, 496−510.
(59) Schwartz, T.; Behlke, J.; Lowenhaupt, K.; Heinemann, U.; Rich, A. *Nat. Struct. Biol.* **2001**, *8*, 761−765.

scaffold, and of solutions corresponding to peptides with fewer than 20 residues, resulted in 226 proteins. Examination of the structures showed that, in some cases, the motifs involved some residues that are not well exposed at the surface of the protein. Therefore, we systematically calculated the accessible surface area for the residues composing the identified motif. Only the solutions with motifs having an accessible surface area above 5 Å^2 per residue were selected. In addition, we observed that, in certain solutions, the orientation of some Cα−Cβ vectors was significantly different from the corresponding one in the FSKY motif. This is likely due to the fact that the motif sought is composed of Cα and Cβ atoms that are close together. The angle between the Cα−Cβ vector in the identified motif and the corresponding vector in the FSKY motif of BgK was then calculated after superimposition. When one of these angles was greater than 90°, the solution was rejected. Application of the above criteria resulted in 48 solutions. A final examination of the PDB files led to an additional elimination of several solutions. These included proteins interacting with cofactors, chelating ions, and files lacking coordinates corresponding to a significant part of the protein.

Finally, the search through the full PDB done with the values of the parameters described and filtered according to the above criteria resulted in 20 solutions. These solutions and topological similarity with the FSKY motif, characterized by the RMSDs from the coordinates of the Cα,Cβ atoms after superimposition on the corresponding atoms in the FSKY motif of BgK, are reported in Table 2. The schematic representations of the 3D structures of these 20 scaffolds are shown in Figure 2. This figure demonstrates that, despite the large diversity of the folds of the protein scaffolds identified by the systematic search through the PDB, the topologies of the Cα,Cβ motifs identified are very similar to that of the FSKY motif in BgK. In this protein, the SKY part of the motif is located at the N-terminal side of the second α helix of BgK. Residues topologically equivalent to this SKY motif are located, in some cases, at the N-terminal part of α helices present in some of the scaffolds identified (1CKK, 1EDL, 1UXD, 1A7F, 1ERF, and 1EAY). However, residues topologically equivalent to the SKY motif can be found on other continuous structural elements, like reverse turns or loops (1BNB, 1D6R, 1H5O, 1J75, 1OQE, 1BOE, 1HZ6, 1J9E, 2PDD, 1F7E, and 1FGP). In the case of the scaffold 1SH1, the SKY equivalent motif is split, as residues topologically equivalent to S23 and Y26 are located in the reverse turn in the N-terminal position of strand 2 of the β sheet and the residue topologically equivalent to K25 is 16 amino acids from the residue topologically equivalent to Y26 and is located in the C-terminal position at the end of strand 3 of the β sheet. In 1SH1, the residue topologically equivalent to F6 is located at the N-terminal position of strand 1 of the β sheet.

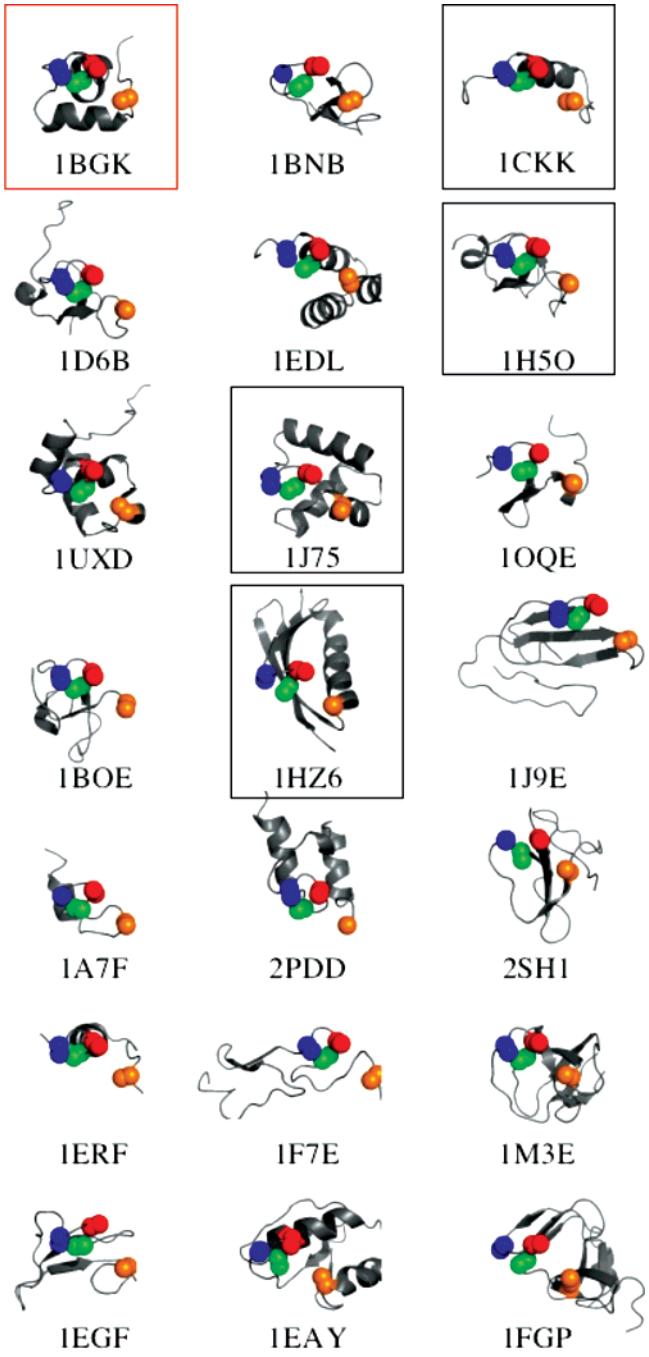


Figure 2. Schematic representation of the 20 scaffolds identified by a systematic search through the PDB of proteins possessing a Cα,Cβ motif topologically equivalent to the Cα,Cβ of the FSKY motif in BgK. The red frame indicates the schematic structure of BgK. Black frames indicate structures of the four scaffolds selected for further experimental evaluation. The main chains of the proteins are shown in ribbon representation. The Cα,Cβ atoms of the FSKY motif or the residues topologically equivalent to those of the FSKY motif are shown in sphere representation in orange (F), blue (S), red (K), and green (Y).

(60) Liu, Y.; Hong, X.; Kappler, J.; Jiang, L.; Zhang, R.; Xu, L.; Pan, C. H.; Martin, W. E.; Murphy, R. C.; Shu, H. B.; Dai, S.; Zhang, G. *Nature* **2003**, *423*, 49−56.
(61) Kalus, W.; Zweckstetter, M.; Renner, C.; Sanchez, Y.; Georgescu, J.; Grol, M.; Demuth, D.; Schumacher, R.; Dony, C.; Lang, K.; Holak, T. A. *EMBO J.* **1998**, *7*, 6558−6572.
(62) O'Neill, J. W.; Kim, D. E.; Baker, D.; Zhang, K. Y. *Acta Crystallogr. Sect. D* **2001**, *57*, 480−487.
(63) Artali, R.; Bombieri, G.; Meneghetti, F.; Gilardi, G.; Sadeghi, S. J.; Cavazzini, D.; Rossi, G. L. *Acta Crystallogr. Sect. D* **2002**, *58*, 1787−1792.
(64) Ludvigsen, S.; Olsen, H. B.; Kaarsholm, N. C. *J. Mol. Biol.* **1998**, *279*, 1−7.
(65) Kalia, Y. N.; Brocklehurst, S. M.; Hipps, D. S.; Appella, E.; Sakaguchi, K.; Perham, R. N. *J. Mol. Biol.* **1993**, *230*, 323−341.
(66) Fogh, R. H.; Kem, W. R.; Norton, R. S. *J. Biol. Chem.* **1990**, *265*, 13016−13028.
(67) Gordon, L. M.; Mobley, P. W.; Pilpa, R.; Sherman, M. A.; Waring, A. J. *Biochim. Biophys. Acta* **2002**, *1559*, 96−120.
(68) Kao, Y. H.; Lee, G. F.; Wang, Y.; Starovasnik, M. A.; Kelley, R. F.; Spellman, M. W.; Lerner, L. *Biochemistry* **1999**, *38*, 7097−7110.
(69) Fushman, D. Personal communication.
(70) Montelione, G. T.; Wuthrich, K.; Burgess, A. W.; Nice, E. C.; Wagner, G.; Gibson, K. D.; Scheraga, H. A. *Biochemistry* **1992**, *31*, 236−249.
(71) McEvoy, M. M.; Hausrath, A. C.; Randolph, G. B.; Remington, S. J.; Dahlquist, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7333−7338.
(72) Holliger, P.; Riechmann, L. *Structure* **1997**, *5*, 265−275.

Despite this splitting, the topological similarity between the Cα,Cβ of residues equivalent to the FSKY motif in BgK is excellent, as shown by the rmsd of their coordinates (0.77 Å; Table 2). The case of 1SH1 illustrates the power of this systematic search through the PDB without any assumption regarding the secondary structure of the scaffold.

Electrostatic Filtering. As already noted, the presence of the FSKY motif in the correct topology is not the sole criterion determining the binding to the targeted receptor. The steric complementarity of the scaffold and the target has been addressed via the steric filtering step as described above. Apart from the residues composing the transferred motif (FSKY in our case), other residues naturally present on the scaffolds may establish either favorable or unfavorable interactions with the target. The most important contribution is due to the long-range electrostatic interactions, which can be attractive or repulsive. It has long been recognized that the electrostatic potential plays a central role in the properties of biological macromolecules through direct electrostatic interactions, solvation free energies, and p*K* shifts.[73−75] The electrostatic potential not only is determined by residues present at the surface of the ligand but also results from the complex charge distribution of the molecule. The electrostatic potential of a macromolecule can be calculated with the Poisson−Boltzmann model, which treats the solute as a low-dielectric region containing the charge distribution surrounded by a high-dielectric region containing the solvent. We calculated the electrostatic potential of the different ligands designed after replacement of the residues topologically equivalent to the FSKY of BgK. The electrostatic potential was calculated with the APBS program.[76] The charges of the atoms used in the calculations were those of the PARSE parameter set.[77] The APBS program calculates the electrostatic potential on a 3D grid encompassing the given protein. Visual inspection of the resulting potential might give insight into the similarity of the potentials of the different scaffolds. However, we performed a semiquantitative comparison of the potentials with the PIPSA program.[78] Using a similarity index, this program systematically compares the grid of electrostatic potential of the scaffold with that calculated for the reference ligand, BgK. This was done for the 20 scaffolds identified in the previous step of the approach using the Hodgkin index.[79,80] By definition, when the electrostatic potentials of the two compared structures are equal, the value of the index is 1; when they are anticorrelated it is equal to −1, and when they are uncorrelated it is 0. Table 3 reports the values of the Hodgkin index corresponding to the electrostatic potential of the entire grid (ES_all) and the value of this index calculated for the electrostatic potential restricted to the binding region (ES_site; see Experimental Section). This region is defined by a cone centered on a vector pointing from the center of gravity of the protein to the center of the amino acids at the interface with the target. The angular extent of the cone is then set to encompass the binding region. In addition, the PIPSA program calculates a shape similarity index that characterizes the shape similarity of a given scaffold with that of BgK either for the full protein (SS_all) or restricted to the target-binding region (SS_site). This index ranges from 0 when no intersection is found between the skins of the compared proteins to 1 when the shapes of the skins are similar. The average value of SS_all calculated for the 20 scaffolds is 0.39 (standard deviation from the mean = 0.05). This reflects the structural diversity of the scaffold identified by the systematic search in the PDB and the fact that no constraints were applied to the overall scaffold shapes. However, the average value of SS_site increases to 0.52, indicating a more pronounced shape similarity in this region as compared to the global one. This is partly due to the steric filtering step that must select proteins having a certain degree of shape complementarity with the target. Most of the values are around 0.5. This might be explained by the facts that, at this step, the structures are treated as rigid and that there is no relaxation of the interface residues between the ligand and the target. However, some scaffolds display extreme values of SS_site, like 1F7E_m (SS_site = 0.1), indicating a very different shape in the binding region as compared to BgK, or 1UXD_m (SS_site = 0.98), suggesting a very similar shape as compared to BgK in this region. The main purpose of using the similarity index was to compare the electrostatic potentials of the 20 proteins identified. The values obtained for ES_all display substantial dispersion (standard deviation from the mean = 0.49). Some of these proteins give an overall potential that is very similar to that of BgK (1BNB_m, 1CKK_5m, 1H5O_m, 1OQE_m, and 1J9E_m), with ES_all values ranging from 0.73

Table 3. Hodgkin Similarity Indexes for the Different Scaffolds Identified As Compared to the Reference Ligand (BgK)[a]

	ES_all	SS_all	ES_site	SS_site
1BGK	0.92	0.73	0.97	0.69
1BNB_m	0.81	0.44	0.93	0.54
1CKK_m	0.75	0.39	0.89	0.48
1D6B_m	0.56	0.41	0.66	0.48
1EDL_m	−0.14	0.43	0.58	0.56
1H5O_m	0.82	0.38	0.94	0.62
1UXD_m	0.86	0.37	0.93	0.98
1J75_m	0.68	0.44	0.72	0.46
1OQE_m	0.73	0.38	0.89	0.51
1BOE_m	0.38	0.39	0.73	0.57
1HZ6_m	0.15	0.39	0.31	0.43
1J9E_m	0.82	0.32	0.94	0.48
1A7F_m	0.59	0.40	0.81	0.57
2PDD_m	0.64	0.35	0.78	0.37
2SH1_m	0.26	0.43	0.40	0.64
1ERF_m	0.66	0.45	0.81	0.55
1F7E_m	−0.37	0.35	0.02	0.10
1M3A_m	−0.09	0.46	−0.27	0.65
1EGF_m	−0.01	0.36	0.05	0.55
1EAY_m	−0.76	0.35	−0.81	0.48
1FGP_m	−0.44	0.27	−0.15	0.47
S_average	0.34	0.39	0.51	0.52
S_sigma	0.49	−0.05	0.50	0.16

[a] ES_all and ES_site are the Hodgkin similarity indexes calculated for the electrostatic potential respectively on the full grid and restricted to the binding site. SS_all and SS_site are the shape similarity indexes calculated respectively for the full protein and the binding site region. These values were calculated after replacement of the FSKY motif of each scaffold indicated by their PDB code and an "m". The mean value (S_average) and the standard error of the m (S_sigma) for the similarity parameters were calculated omitting the 1BGK values.

(73) Smith, P. E.; Van Gunsteren, W. F. In *Computer Simulation of Biomolecular Sytems. Theoretical and Experimental Applications*; Van Gunteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; ESCOM: Leiden, 1993; pp 182−212.
(74) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(75) Gilson, M. K. *Curr. Biol.* **1995**, *5*, 216−223.
(76) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(77) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(78) Blomberg, N.; Gabdoulline, R. R.; Nilges, M.; Wade, R. C. *Proteins: Struct., Funct. Genet.* **1999**, *37*, 379−387.
(79) Hogkin, E. E.; Richards, W. G. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1987**, *14*, 105−110.
(80) Wade, R. C.; Gabdoulline, R. R.; De Rienzo, F. *Int. J. Quantum Chem. Quantum Biol. Symp.* **2001**, *83*, 122−127.

Figure 3. Amino acid sequences of the proteins produced in this study. Recombinant proteins 1J75, 1J75_m, and its mutants are obtained by thrombin cleavage of $(His)_6$-tagged proteins and are composed of four N-terminal residues resulting from the DNA construct (shaded in gray) and residues 8−70 of 1J75. Mutations corresponding to the FSKY motif's residues are shaded in red or in green when already present in the native sequence. Other mutations are shaded in cyan.

to 0.86 (Table 3). However, some scaffolds produce anticorrelated electrostatic potentials as compared to that of BgK (1EAY_m and 1FGP_m), while some are weakly correlated or uncorrelated (1HZ6_m, 1EGF_m). The comparison of the ES_site values leads to a similar conclusion. Some scaffolds give ES_site values close to 1 (1BNB_m, 1CKK_m, 1H5O_m, 1J9E_m, 1ERF_m, and 1A79_m), suggesting very similar electrostatic potentials as compared to BgK in the binding region. Scaffold 1EAY_m displays an anticorrelated potential as compared to BgK (ES_site = −0.81), while some other scaffolds display uncorrelated potentials as compared to the reference ligand. It should be noted that the values calculated for ES_all and ES_site display good correlation. This may reflect the facts that the electrostatic potential is determined by the complete charge distribution and that the electric field has a long-range effect. However, 1EDL_m is an exception to this correlation as, despite the fact that the overall potential of the protein is not correlated with that of BgK, the electrostatic potentials in the binding region of the two proteins are correlated (ES_site = 0.58). The results of the comparison of the different electrostatic potentials of the 20 scaffolds identified by our systematic topological search through the PDB allow us to identify a set of 11 scaffolds that display a potential in the binding region similar to that of the reference ligand (ES_site greater than 0.7; 1BNB_m, 1CKK_m, 1H5O_m, 1UXD_m, 1J75_m, 1OQE_m, 1J9E_m1A7F_m, 2PDD_m, and 1ERF_m).

Structural Characterization of the Transferred Motif. In order to evaluate the capacity of our design strategy to obtain ligands of the Kv1.2 channel, we selected some of these scaffolds for experimental investigations. 1J75_m, several mutants of this protein, and the corresponding native protein 1J75 were produced by molecular biology as described in the Experimental Section. The synthesis of several scaffolds was undertaken by peptide chemical synthesis. Some of these scaffolds were successfully synthesized (1CKK_m, 1H5O_m, and 1HZ6_m), purified, and characterized (Experimental Section). The latter scaffold, which displayed an electrostatic potential weakly correlated with that of the reference ligand BgK, was selected in order to evaluate the impact of the electrostatic potential on the binding to the target. We also produced the native scaffolds (1J75, 1CKK, 1H5O, and 1HZ6) to control the influence of the replacement of the residues topologically equivalent to the FSKY motif in BgK on the structure of these scaffolds that is central in the current approach. The dichroic spectra of the different scaffolds after transfer of the FSKY motif were compared to those of the corresponding native scaffolds (see Experimental Section and Supporting Information). No significant differences were observed in these cases, suggesting that the replacement of the residues topologically equivalent to those of the FSKY motif did not alter the overall 3D structure of these scaffolds. However, to better characterize the topology of the FSKY motif in 1J75_m, the 3D structure of this scaffold was determined by X-ray diffraction. Several mutants of 1J75_m were produced for further characterization of the binding of 1J75_m to the Kv1.2 target. Co-crystallization of one of these mutants, 1J75_m_K130A, produced with a high yield in our expression system, with a seven-base-pair DNA duplex under conditions similar to those reported in the original study[59] (Experimental Section) led to

Table 4. Summary of Data Collection and Refinement Statistics for the Structure of 1J75_m_K130A.

Data Collection	
crystal	prismatic
space group	*P*61
unit cell dimensions (Å)	$a = b = 79.91$, $c = 55.33$
unit cell angles (°)	$\alpha = \beta = 90.00$, $\gamma = 120.00$
resolution (Å)	65−1.7
measured/unique reflections	155 589/22 226
completeness (%)	99.9 (100.0)
R_{merge} (%)	8.2 (41.7)
$I/\sigma(I)$ ratio	16.1 (3.1)
multiplicity	7.0 (7.0)
Wilson *B* factor (Å^2)	25.8
Refinement Parameters	
no. of atoms	
protein chain A	477
protein chain D	459
DNA chain B	124
DNA chain E	124
water	228
average overall *B* factors	
protein chain A	31.9
protein chain D	32.5
DNA chain B	30.0
DNA chain E	32.1
water	41.8
R factor (%)/R_{free} (%)	24.1/30.5
RMS_bond (Å)	0.013
RMS_angle (Å)	2.007
RMS_chiral (Å)	0.158
figure of merit	0.763

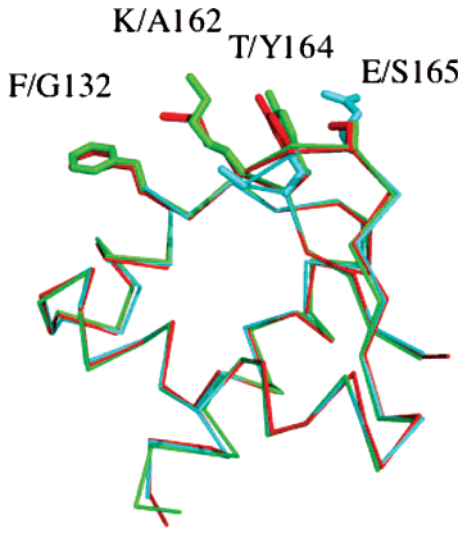


Figure 4. Superimposition of the structures of the two proteins 1J75_m present in the asymmetric unit (1J75_m_K130A_A in red and 1J75_m_K130A_D in green) with the structure 1J75 (in cyan). These structures were superimposed by minimizing the root-mean-square deviation from the positions of the Cα atoms of residues 113−169. The structures are shown in ribbon representation. The side chains corresponding to the position of the FSKY motif are shown in stick representation. The hydrogens are omitted.

Table 5. Comparison of the Structures of 1J75_m and 1J75[a]

		1J75		BgK
	all Cα	all Cα,Cβ	motif Cα,Cβ	motif Cα,Cβ
1J75_m_K130A_A	0.34	0.45	0.59	0.88
1J75_m_K130A_D	0.49	0.59	0.54	0.84
1J75				0.35

[a] 1J75_m_K130A_A and 1J75_m_K130A_D represent the two monomers found in the asymmetric unit of 1J75_m_K130A. "all" indicates superimposition involving all residues of the protein; "motif" indicates superimposition involving the residues corresponding to either the FSKY motif or the native one (1J75).

two crystal forms. One of these corresponds to that obtained in the original study of the structure 1J75[59] but diffracted at a maximum resolution of 2.8 Å. We obtained a second crystal form that diffracted to a maximum resolution of 1.7 Å. The X-ray diffraction study of this latter crystal allowed us to determine the 3D structure of 1J75_m_K130A by molecular replacement using the structure 1J75. The data collection and refinement parameters of this second form are shown in Table 4. Although they have different space groups, the overall structure of 1J75_m_K130A is similar to that of 1J75 previously reported. The asymmetric unit contains two monomers of 1J75_m_K130A (chains A and D) and two strands of left-handed Z-DNA (chains B and E). The amino acid interactions with the DNA are similar to those observed in the original structure, 1J75. Figure 4 shows the superimposition of the structure 1J75 with that of 1J75_m_K130A determined in the present study after minimization of the rmsd of the coordinates of Cα carbons between the two proteins (Table 5). This shows that the overall structures of the two monomers of 1J75_m_K130A are very similar to that of 1J75 (rmsd's between the Cα coordinates (residues 113−169) of 1J75 and those of 1J75_m_K130A_A and 1J75_m_K130A_D are 0.34 and 0.49, respectively). A similar observation is made when calculating the rmsd's between the Cα,Cβ coordinates from residues 113 to 169 (rmsd = 0.45 and 0.59 Å for 1J75_m_K130A_A and 1J75_m_K130A_D,

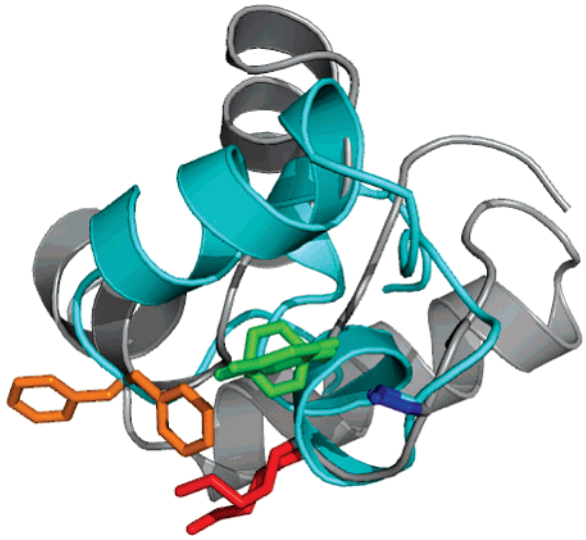


Figure 5. Superimposition of the structure of 1J75_m_K130A_A on that of 1BGK after minimization of the root-mean-square deviation from the coordinates of the Cα and Cβ atoms of the residues topologically equivalent to those in the FSKY motif in 1J75_m and these residues in BgK. The backbones of the proteins are shown in cartoon representation in gray for 1J75_m and in cyan for BgK. The side chain of the FSKY motif is shown in stick representation in orange (F), blue (S), red (K), and green (Y). The hydrogen atoms are omitted.

respectively). The rmsd's of the Cα,Cβ coordinates of the residues corresponding to the FSKY motif between 1J75_m and 1J75 demonstrate a very high topological similarity: the rmsd of the Cα,Cβ coordinates of the four residues of the motif as compared to the corresponding one in 1J75 are equal to 0.59 and 0.54 Å for 1J75_m_K130A_A and 1J75_m_K130A_D, respectively. Comparison of the rmsd of Cα,Cβ of the FSKY motif in BgK with that observed in the structure of 1J75_m_K130A shows that this scaffold reproduces well the topology of the reference ligand (rmsd = 0.88 and 0.84 Å for 1J75_m_K130A_A and 1J75_m_K130A_D, respectively). Examination of the superimposition of the structure of 1J75_m_K130A with that of 1BGK (Figure 5) shows that the orientations of the side chains of the SKY residues in 1J75_m_K130A are similar to those observed in the reference structure of BgK. The F132 side chain of 1J75_m_K130A has a different orientation in the crystal (*t* conformation, $\chi_1 = 180°$) than that observed for F6 in the structure of BgK (g_+ conformation, $\chi_1 = -60°$). However, there is no obvious reason to exclude the possibility that the F132 side chain could adopt a g_+ conformation similar to that of F6 in BgK when interacting with the Kv1 channels.

Binding Activities of the Designed Protein Ligands. The Kv1 binding activities of the proteins designed with our approach were assayed by competition experiments. Proteins IJ75_m, 1CKK_m, and 1H5O_m inhibit the binding of ^{125}I-BgK_W5Y,Y26F to membranes from cells producing Kv1.2 channels with K_i constants respectively equal to 1.5, 1.6, and 0.5 μM (Table 6). The results of these competition experiments demonstrate that these three ligands are able to bind the Kv1.2 channel at the same site as BgK. In addition, the values of K_i observed for the native proteins 1J75, 1H5O, and 1CKK give interesting information regarding the process of selection of these scaffolds. Indeed, the inhibition constants for the native scaffold are in the range 1−30 μM. The most straightforward interpretation of this observation is that these proteins were selected for their steric and electrostatic similarity with the

Table 6. Dissociation Constants of the Different Protein Ligands Designed in This Study To Bind to the Kv1.2 Potassium Channel (Inhibition of ^{125}I-BgK(W5Y/Y26F) of Membranes from TsA-201 Cells Expressing Human Kv1.2

	Kv1.2	
	K_i (μM)	Hn[a]
1J75_m	1.5 ± 0.7	1.2 ± 0.4 (10)
1J75	26 ± 9.4	0.9 ± 0.1 (7)
1J75_m_K130A	2.9 ± 1.4	1.1 ± 0.5 (5)
1J75m_K117A	1.7 ± 0.4	0.7 ± 0.2 (3)
1J75_m_K164A	6.4 ± 1.8	1.1 ± 0.4 (3)
1J75_m_F132A	3.8 ± 1.1	0.9 ± 0.4 (3)
1J75_m_C138V	3.8 ± 0.4	0.9 ± 0.2 (4)
1J75_m_FSKY/A	17.2 ± 6.0	0.8 ± 0.2 (4)
1H5O_m	0.5 ± 0.1	0.9 ± 0.2 (3)
1H5O	1.1 ± 0.4	0.9 ± 0.1 (3)
1H5O_m_K33A,Y34A	6.5 ± 3.2	1.0 ± 0.2 (3)
1H5O_m_K33E	7.5 ± 4.0	1.5 ± 0.4 (3)
1CKK_m	1.6 ± 0.5	1.3 ± 0.6 (3)
1CKK	2.3 ± 0.5	1.2 ± 0.2 (3)
1HZ6_m	>100	
1HZ6	>100	

[a] Hill numbers. The number of experiments are given in parentheses. The proteins designed are indicated by their original PDB code with "_m" indicating the replacement of the residues topologically equivalent to the FSKY motif of BgK by the corresponding residues.

reference ligand, BgK. Therefore, assuming a similar binding mode of the scaffold in the presence or absence of the FSKY motif, the affinity observed may reflect the global properties of these scaffolds. As compared to the native scaffolds, the presence of the FSKY motif increases the affinity for the Kv1.2 channel by a factor ranging from about 2 (1CKK_m) to about 20 in the case of 1J75_m. These observations suggest that the FSKY motif in our designed ligand plays a role in the interaction with the target. The 1HZ6_m protein did not exhibit any inhibition of the binding of ^{125}I-BgK_W5Y,Y26F at a concentration of 100 μM. In order to characterize the interaction of 1J75_m and 1H5O_m better, we produced several mutants of these ligands. In the case of 1H5O_m, the residues of the KY dyad were mutated into alanine (1H5O_m_K33A,Y34A). This mutant displays an approximately 6-fold reduction in affinity for the Kv1.2 channel as compared to 1H5O_m (Table 6). A similar observation was made with a second mutant of 1H5O_m (1H5O_m_K33E), where the K residue was replaced with a glutamate. This mutant displays a roughly 15-fold reduction in affinity for the Kv1.2 channel. In the case of 1J75_m, several mutants were produced. The replacement of the residues of the FSKY motif by alanine residues led to an approximately 10-fold reduction in affinity for the Kv1.2 channel. Single mutants of residues K164 and F132 of the FSKY motif led to 4- and 2-fold reductions in affinity, respectively. Other lysine residues were mutated on the 1J75_m scaffold. Assuming that 1J75_m interacts in the binding mode defined by positioning the FSKY motif at a location similar to that in the complex of BgK with Kv1.2, K117 is located at the opposite side of the protein as compared to the Kv1.2 channel. The results obtained for the mutant 1J75_m_K117A are consistent with this hypothesis, as this mutant exhibits no variation in binding affinity as compared to 1J75_m (Table 6). The mutant 1J75_m_K130A was also produced and its binding affinity determined. While K130 does not participate in the binding motif, it is located on the side of the protein that interacts with the Kv1.2 channel and is close to the binding motif. When mutating the K130 residue into alanine, we observed a 2-fold reduction in affinity compared to 1J75_m. Altogether, these results are consistent with the presumed binding mode with the FSKY motif in a location similar to that of BgK. However, it should be noted that the replacement of the residues composing the functional motif tends to give a lower reduction in the binding affinity of these ligands for the Kv1.2 channel as compared to the results obtained with BgK. This might be due to the fact that, despite the overall conservation of the topology of the functional motif in these scaffolds, fine-tuning of the interactions with the Kv1.2 channel is not achieved. However, caution is warranted in the interpretation of the data from alanine scanning, which perturbs the considered interaction; the observed variations can arise from various sources and are hard to assign to specific atom−atom interactions.[2] Therefore, if alanine scanning identifies residues involved in the interactions, one can hardly expect the quantitative aspects of such replacements to be translated into different structural contexts.

Models for the Interactions of the Designed Protein Ligands with the Kv1.2 Channel. The results obtained for the protein 1HZ6_m led us to investigate in more detail the consequences of positioning the different scaffolds in the presumed binding mode reproducing that of the FSKY motif, as in the model of BgK interacting with the Kv1 channel. We therefore built some models of the interactions of BgK, 1J75_m, 1H5O_m, 1CKK_m, and 1HZ6_m with the S5−S6 part of the Kv1.2 channel. We used, as a starting point, the structure of the Kv1.1/BgK complex.[81] The X-ray structure of Kv1.2 (2A79)[92] was superimposed on that of the S5−S6 part of Kv1.1 in the modeled complex (Kv1.1/BgK). The rmsd calculated for the Cα atoms of the overall structure of this model from that of the X-ray structure 2A79 in the S5−S6 part of the channel was 2.01 Å. The similarity between the two structures is higher in the region where BgK interacts than at the lower part of the channel. This led to an initial structure of the Kv1.2/BgK complex that was further refined by molecular dynamics calculations and energy minimization. A similar strategy was used to build models of 1J75_m, 1H5O_m, 1CKK_m, and 1HZ6_m interacting with Kv1.2. Initial complexes of these structures with the S5−S6 part of the Kv1.2 channel were obtained by superimposing the FSKY motifs in these proteins on that of BgK in the model BgK/Kv1.2 complex. These initial 3D structures of the ligands/S5−S6 part of the Kv1.2 channel were then refined by molecular dynamics calculations and energy minimization under positional restraints (see Experimental Section). The scheme applied allowed global motions of the scaffold relative to the Kv1.2 channel and all relaxation of the side chains of the ligand and the target and ensured preservation of its overall 3D structure while maintaining the position of the FSKY motif as observed in the Kv1.2/BgK model. The results reported below are taken from the last structure of the quenched molecular dynamics protocol (Experimental Section). Table 7 reports a set of parameters obtained from analysis of these models for 1BGK, 1J75_m, 1H5O_m, 1CKK_m, and 1HZ6_m. The rmsd from the Cα atoms of the ligands was obtained after superimposition of the structure in the complex with the initial structure as reported in the PDB (1BGK, 1J75, 1H5O, 1CKK, and 1HZ6). These values range from 0.37 Å for 1J75_m to 1.22 Å for 1CKK_m, showing that the overall structures of the ligands were preserved during the

(81) Gilquin, B.; Braud, S.; Eriksson, M. A. L.; Roux, B.; Bailey, T. D.; Priest, B. T.; Garcia M. L.; Menez, A.; Gasparini, S. *J. Biol. Chem.* **2005**, *280*, 27093−27102.

Table 7. Summary of the Parameters Calculated in the Models of Interaction of the S5−S6 Part of the Kv1.2 Channel with BgK and the Four Proteins Reported in This Study (1J75_m, 1H5O_m, 1CKK_m, and 1HZ6_m)[a]

	protein ID				
	1BGK	1J75_m	1H5O_m	1CKK_m	1HZ6_m
RMSD_Cα (Å)	0.55	0.37	0.97	1.22	0.53
RMSD_motif (Å)	0.36	0.53	0.65	0.44	0.82
ΔASA_interf_tot (Å^2)	1583	1989	1384	1856	2031
ΔASA_interf_lig (Å^2)	782	889	703	991	964
Nb_Res_Inter	16	24	15	16	23
% Res_NP	37.5	50	46.6	62.5	43.5
Gap Vol (Å^3)	1290	1704	1326	1635	1872
Gap Vol Index (Å)	1.65	1.92	1.89	1.70	1.89
Nb HB	10	11	9	5	9
Nb HB/100 Å^2 Lig	1.3	1.2	1.3	0.5	0.9
Qtot ligand	5	5	8	6	−1
Qtot interf lig	5	3	5	3	−1
Nb SB	3	5	3	3	4
Nb g+ SB	3	3	3	3	1
Nb g- SB	0	2	0	0	3
Nb Net Ch Inter	5	7	7	5	11
Nb Net Ch Uncomp	2	2	4	2	7

[a] RMSD_Cα, root-mean-square deviation from the coordinates of the Cα atoms of the considered protein after refinement as compared to the initial one after superimposition of the two structures on these atoms; RMSD_motif, root-mean-square deviation from the coordinates of the non-hydrogen atoms of the sidechains of the residues composing the FSKY motif using the initial structure of BgK in the KV1.2/BgK complex as reference; ΔASA_interf_tot, variation in accessible surface area of the ligand and protein upon binding; ΔASA_interf_lig, variation in accessible surface area of the ligand upon binding; Nb_Res_Inter, number of residues of the ligand present at the interface; % Res_NP, percentage of nonpolar residues of the ligand at the interface; Gap Vol, gap volume at the interface; Gap Vol Index, gap volume index; Nb HB, number of hydrogen bonds at the interface; Nb HB/100 Å^2 Lig, number of hydrogen bonds per 100 Å^2 of interface; Qtot ligand, global charge of the ligand; Qtot Interf Lig, total charge of the residues of the ligand present at the interface; Nb SB, number of intermolecular salt bridges; Nb g+ SB, number of groups of the ligand bearing a net +1 charge and involved in intermolecular salt bridges; Nb g- SB, number of groups of the ligand bearing a net −1 charge and involved in intermolecular salt bridges; Nb Net Ch Inter, number of groups of the ligand bearing a net +1 or −1 charge located at the interface; Nb Net Ch Uncomp, number of groups of the ligand bearing a net +1 or −1 charge not involved in salt bridges.

refinement. The rmsd from the position of the non-Cα heavy atoms of the FSKY motif was calculated using the initial structures of these complexes prior to refinement as reference. These values ranged from 0.36 Å for BgK to 0.82 Å for 1HZ6_m, demonstrating that the refinement protocol conserved the position of the FSKY motif.

A number of global parameters calculated in these models indicate the consequences of this positioning in terms of interactions involving other parts of the proteins. The size of the interface per monomer, evaluated with the ΔASA parameters, shows that in all cases it corresponds to the expected values for protein−protein interactions (from 703 Å^2 for 1H5O_m to 991 Å^2 for 1CKK_m), as reported in the analysis of natural hetero-complexes in the protein database.[10,27,29,38,82] One very important parameter regarding protein−protein interactions is surface complementarity. This can be evaluated by calculating the gap volume between the ligand and the target. These values calculated for the five complexes range from 1290 Å^3 for 1BGK to 1872 Å^3 for 1HZ6_m. However, the gap volume depends on the size of the interface, and a gap volume index has been proposed for relevant comparison of the surface complementarity of several complexes.[40] The gap volume index is defined as the ratio of the gap volume to the size of the interface. In the present case, the gap volume index calculated for 1BGK is 1.65 Å. In the case of 1CKK_m, the gap volume is close to the value calculated for 1BGK (1.70 Å). The gap volume indexes calculated for 1J75_m, 1H5O_m, and 1HZ6_m are greater (about 1.9 Å), demonstrating a slightly lower surface complementarity with the target for these proteins in the models. However, all values calculated in the fives models remain in the range observed at the interface of hetero-complexes analyzed in the protein database.[29]

The models of complexes of the S5−S6 part of Kv1.2 with the five proteins allow us to analyze the intermolecular interactions that can take place in the presumed binding mode. We calculated the number of hydrogen bonds between the Kv1.2 channel and the different ligands in the derived models. The criteria for assignment of hydrogen bonds were those used in the HBPLUS program[83,84] (Experimental Section). The number of hydrogen bonds ranges from 5 in 1CKK_m to 11 in 1BGK. These hydrogen bonds involve donors and acceptors belonging to either the main chain or the side chains in the ligand and the target. However, to gain insight into the interaction, the number of hydrogen bonds must be compared to the size of the interface. To this end, it has been proposed to use the surface density of hydrogen bonds, calculated as the number of bonds per 100 Å^2 of interface of the considered protein. The values of this parameter calculated for the five models are reported in Table 7. Higher values are observed for 1BGK, the reference ligand, 1H5O_m (1.3 HB/100 Å^2), and 1J75_m (1.2 HB/100 Å^2), while lower values are found for 1HZ6_m and 1CKK_m (0.9 and 0.5 HB/100 Å,^{2} respectively). The surface densities of hydrogen bonds, calculated with the same geometric criteria as in the present report for the five models, are in the range observed at the interface of hetero-complexes taken from the protein database.[9,29,82]

As already pointed out, electrostatic interactions play an important role in intermolecular interactions. It has recently been shown that the electrostatic binding free energy is insensitive to the total molecular charge of the ligand in the case of Barstar.[85] Assuming unshifted p*K*'s at neutral pH, the total molecular charge of the designed protein in the present work ranges from −1 for 1HZ6_m to 8 in the case of 1H5O_m (Table 7). The net charge for the residues at the interface with the target is 5 for 1BGK and 1H5O_m, and 3 for 1J75_m and 1CKK_m. The net charge of the interface of 1HZ6_m is −1. This may be related to lesser electrostatic similarity of this ligand as compared to BGK. The electrostatic potentials at the surfaces of the proteins 1BGK, 1J75_m, 1H50_m, 1CKK_m, and 1HZ6 are shown in Figure 6. The region of the FSKY motif is indicated with a circle. It illustrates the electrostatic similarity already pointed out for 1BGK, 1J75_m, 1H5O_m, and 1CKK_m. In the case of 1HZ6_m, the electrostatic potential at the interaction surface appears significantly different from that observed for the four other proteins. This difference appears to be more pronounced in regions other than that of the FSKY motif. This observation prompted us to investigate the presence of unfavorable or repulsive interactions between the target and the ligands. In the context of the present work, positioning of the ligand to reproduce the location of the FSKY motif as in the complex

(82) Xu, D.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1997**, *10*, 999−1012.
(83) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97.
(84) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777−793.
(85) Lee, L. P.; Tidor, B. *Protein Sci.* **2001**, *10*, 362−377.

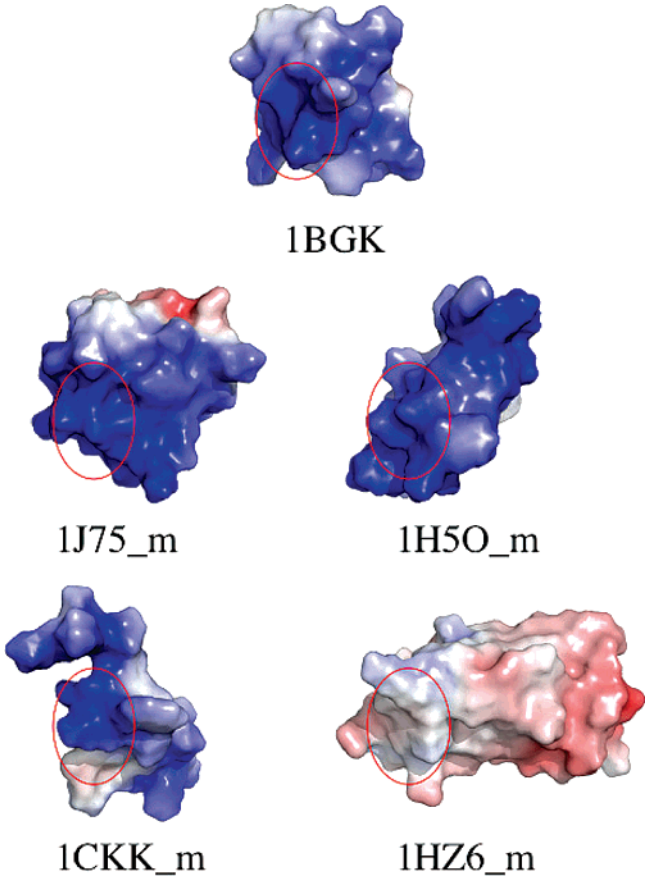


Figure 6. Comparison of the electrostatic potentials of the proteins selected for experimental evaluation and calculated with the Poisson−Boltzmann equation. Only the solvent-accessible surfaces of these proteins are shown. These surfaces are colored according to the values of the electrostatic potential on a scale ranging from −5 k_bT/e (in red) to +5 k_bT/e (in blue).

Kv1.2/BgK may force some residues bearing net charges of identical sign to be close together. Such a situation would be unfavorable to the binding of the considered ligand. No such electrostatic clash was observed in the five models when using a distance criterion of 4.0 Å between groups bearing net charges of identical sign. Another important consequence of the presence of charged groups at the interface is the possible formation of salt bridges between a residue of the target possessing a net charge and a residue of the ligand possessing a net charge of opposite sign. Salt bridges are intuitively thought to contribute to stability. However, a theoretical study showed that salt bridges can be electrostatically destabilizing, as in certain cases the unfavorable desolvation free energy may not compensate for the favorable electrostatic interaction occurring between the charged groups. However, salt bridges can occur at protein−protein interfaces, and it is very important to evaluate the number of uncompensated charges present at the interface. Such uncompensated charges would make an unfavorable contribution to the binding of the ligand via the necessary desolvation penalty.[86,87] The formation of intermolecular salt bridges was systematically analyzed in the five complexes reported here. In addition to the charged groups of residues K, R, E, and D, the N-terminal and C-terminal groups are considered to bear a net charge. An intermolecular salt bridge is considered to be formed when a distance lower than 3.0 Å separates charged groups of opposite sign across the interface. Three intermolecular salt bridges are formed in 1BGK, 1H5O_m, and 1CKK_m_m, four in 1HZ6_m, and five in 1J75_m (Table 7). It should be noted that the net charges of the groups of ligands involved in salt bridges are mainly positive, with the notable exception of 1HZ6_m, which has three negative groups and one positive group involved in salt bridges. Comparison of the number of charged groups of the ligand present at the interface with the number of salt bridges formed sheds light on the interaction (Table 7). In 1BGK, 1J75_m, and 1CKK_m, there are two uncompensated charged groups at the interface, while this number increases to four in 1H5O_m. In the case of 1HZ6_m, they are seven uncompensated charged groups at the interface with the S5−S6 part of the Kv1.2 channel. This observation suggests that positioning the 1HZ6_m scaffold to reproduce the location of the FSKY motif as observed in BgK would bury seven net charges that could not compensate for the unfavorable energetic cost of their desolvation upon binding the Kv1.2 channel. This might explain the absence of binding observed at 100 μM for the 1HZ6_m protein.

Conclusion

The protein ligand design approach reported in the present study and applied to the Kv1.2 potassium channel is based on the transfer of a functional motif of BgK, a natural ligand of this channel. This motif is used with two goals in mind: (i) to reproduce the interactions that arise between the reference ligand and the target via these residues and (ii) to provide a shape guide for screening the PDB to identify scaffolds that are sterically and electrostatically compatible for the interaction with the target and that are able to reproduce the functional topology of the binding motif. Our approach contrasts with the previously reported methods of protein ligand design that are generally based on the transfer of a group of functionally important residues on a structural element similar to that bearing the motif in a natural ligand of the considered target. The standard approach considerably limits the diversity of scaffolds that can be used to target a particular receptor. In the present work, no constraint is applied to the secondary structure of the scaffolds. This is practically possible because we developed a program that systematically searches, through the very large number of protein structures accessible in the PDB, for a motif of residues in a topology similar to that of the functional motif to be transferred. While important successes have been achieved with the standard approach, it should be noted that the affinity obtained for the reported mini-proteins prior to optimization is often in the range 10−100 μM or more.[14,15,17] The in silico approach reported in the present study led to 20 proteins that reproduce the topology of the functional motif of BgK and are sterically compatible for the interaction with the Kv1.2 channel. These proteins have secondary structures unrelated to that of the natural ligand BgK. For experimental investigations, three structures electrostatically similar to BgK in the binding region were produced. These three ligands bind the Kv1.2 channel with micromolar or submicromolar affinity at the same site as the reference ligand BgK. The high yield of the process may be due to the fact that the overall properties of the scaffold were taken into account explicitly in the design of these protein ligands. The most potent ligand displayed an affinity of 500 nM without further optimization. Even though the affinity range of the designed ligand appears significantly lower than that of

(86) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1983**, *168*, 867−885.
(87) Rashin, A. A.; Honig, B. *J. Mol. Biol.* **1984**, *173*, 515−521.

BgK (about 50 pM), it should be kept in mind that a difference of about 10^4 in affinity corresponds to only about 5.5 kcal/mol in free energy, which corresponds to a small number of interactions. Optimization of the protein ligands reported here could take advantage of observations made in the molecular models to evaluate the consequences of the presumed binding mode in terms of intermolecular interaction. Improving the affinity of the present ligands could be envisaged by electrostatic optimization to reduce the number of uncompensated polar groups. Such an approach has been reported in the case of the stabilization of the Arc repressor and led to an increase in stability of about 5 kcal/mol.[88] Other structure-based or random approaches can be envisaged as well to increase the affinity of the designed protein ligands. The first application of this novel structure-based secondary-structure-independent approach—targeting the Kv1.2 channel—gave very encouraging results. This work demonstrated that it is possible to design protein ligands by transferring a binding motif in its functional topology irrespective of the structure of the scaffold selected to reproduce this functional topology. This approach will likely be further evaluated and improved through application to the design of ligands targeting other receptors or enzymes. The evaluation of the influence of the size and of the particular topology of the motif to transfer on the number of solutions returned will also be necessary to better estimate the potential of this approach. Improving the steric filter to select the scaffolds could also permit researchers to better take advantage of the considerable repertoire of the PDB to design protein ligands.

Experimental Section

Structural Data. The structure of the motif to be transferred was derived from the NMR structure 1BGK deposited in the PDB. An average structure was calculated by averaging the coordinates of the atoms for the 15 different conformers deposited in the PDB. The resulting average structure was then minimized with version 27 of the CHARMM program.[89] We used CHARMM force field version 22.[90] Energy minimization with five cycles of 1000 steps, using the Adopted Basis Newton−Raphson method, was carried out with harmonic restraints on the position of the heavy atoms (C,N, O, and S). In the five cycles of energy minimization, the harmonic constants applied to the backbone heavy atoms were varied from 100 to 20 kcal mol^{-1} A^{-2} in steps of 20 kcal mol^{-1}, and harmonic constants used for the side-chain heavy atom were varied from 10 to 2 kcal mol^{-1} A^{-2} in steps of 2 kcal mol^{-1} A^{-2}. The resulting structure was submitted to a final 10 000-step energy minimization with positional restraints and harmonic constants for the backbone heavy atoms, and the side chains were respectively set to 5 and 0.5 kcal mol^{-1} A^{-2}. The resulting average structure of BgK was superimposed on the 15 different conformers by minimizing the rmsd from the positions of the 8 Cα and Cβ carbons of the motif. For all conformers, the resulting rmsd was lower than 0.42 Å.

Electrostatic Potential Calculations. The electrostatic potentials of the different scaffolds reported in this study were calculated using the Poisson−Boltzmann model with the APBS program.[76] The coordinates of the structures used in these calculations for the different scaffolds were those obtained after superimposition of the FSKY motifs of each scaffold on the motif of BgK The charge and radius were those of the PARSE parameter set.[77] Residues taken into account were those present in the original PDB files. The grid dimensions and their spacings were defined in order to encompass all the scaffolds studied. This allowed further comparison of the electrostatic potential using the PIPSA software.[78] The electrostatic potential was calculated using the focusing method. The nonlinear version of the Poisson−Boltzmann equation was used with an ionic strength of 150 mM, defined as one ion of charge +1 at a concentration of 0.075 mM and one ion of charge −1 at the same concentration, the radius of these ions being set to 2.0 Å. The dimensions of a coarse grid in the three directions were set to 97 points with a spacing of 2.0 Å. The solvent radius was set to 1.4 Å. The solute and solvent dielectric constants were set respectively to 1 and 80. During the first calculation, we used the multiple Debye−Hückel boundary condition. During the focusing step, the grid dimensions were set to 97 points in each direction, the spacing was set to 0.8 Å, and the boundary condition was obtained from the coarse grid calculation. The electrostatic potentials calculated were written in the UHBD format to allow further analysis with the PIPSA program. The electrostatic potential was compared using version 2 of the PIPSA program. The skin parameter, corresponding to the thickness of the electrostatic grid region analyzed, was set to 4 Å. This region was defined at a distance of 3 Å from the surface of the proteins, the probe radius. The similarity of the electrostatic potentials and shapes between the two proteins was calculated at the intersection of their so-called skins with the Hodgkin similarity indexes.[79,80]

Molecular Modeling. (a) Modeling of the Proteins/S5−S6 Part of the Kv1.2 Complexes. The structures of the complexes between the S5−S6 part of the Kv1.2 channel and BgK or the four proteins were built using, as starting point, a model of the Kv1.1/BgK complex obtained with a docking procedure based on back-calculated coupling free energy obtained from double mutant cycles.[81] The X-ray structure of the S5−S6 domain of the Kv1.2 channel[84] was superimposed on the corresponding part of the Kv1.1 in the Kv1.1/BgK model. The structure of the BgK/Kv1.2 complex was then relaxed using a protocol of molecular dynamics and energy minimization with positional restraints using version 27 of the CHARMM program.[89] We used CHARMM force field version 22.[90] In this protocol, positional restraints were applied to the Cα atoms of the S5−S6 part of the Kv1.2 channel. The force constant used for the Cα was 50 kcal mol^{-1} $Å^{-2}$ for residues located at a distance greater than 7 Å from the ligand and 5 kcal mol^{-1} $Å^{-2}$ for Cα of residues located at a distance less than 7 Å. For the non-hydrogen atoms of the target (other than Cα), a force constant of 5 kcal mol^{-1} $Å^{-2}$ was applied for residues located at a distance greater than 7 Å from the ligand, while harmonic positional restraints with a force constant of 0.005 kcal mol^{-1} $Å^{-2}$ were applied to the non-hydrogen atoms other than the Cα of residues located at a distance lower than 7 Å. For the ligand, no positional restraints were applied except to the heavy atoms of the side chain of the FSKY motif, using the coordinates of the FSKY motif in the BgK/Kv1.2 complex as reference. For these restraints, a force constant of 5 kcal mol^{-1} $Å^{-2}$ was applied during the refinement. In addition, a set of intramolecular Cα−Cα distance restraints was applied to the structure of the ligand. This scheme allowed global motions of the scaffold relative to the Kv1.2 channel and all relaxation of the side chains of the ligand and target, and ensured the preservation of its overall 3D structure while maintaining the position of the FSKY motif as observed in the Kv1.2/BgK model. The first step of the protocol involved the electrical neutralization of the complex with sodium or chloride ions, depending on the initial charge of the complex. These counterions were successively added in positions corresponding to the energy minimum of the system until its global charge reached 0. The five complexes were submitted to 750 000 steps of molecular dynamics calculation at 300 K using the Verlet algorithm with an integration time step of 0.0005 ps. Thirty structures, each taking 25 000 steps, were submitted to 2500 steps of energy

(88) Hendsch, Z. S.; Jonsson, T.; Sauer, R. T.; Tidor, B. *Biochemistry* **1996**, *35*, 7621−7625.
(89) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(90) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(91) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Computer Phys. Commun.* **1991**, *62*, 187−197.
(92) Long, S. B.; Campbell, E. P.; MacKinnon, R. *Science* **2005**, *309*, 897−903.

minimization and analyzed. For each complex, the 30 structures were analyzed to check for convergence of the results. During these calculations, the nonbonded interactions were modeled using a Lennard-Jones function and a Coulombic electrostatic term with a nonbonded cutoff of 15 Å. The dielectric constant was set to 1. All calculations were carried out in a vacuum.

(b) Analysis of the Protein/Kv1.2 Interactions. The models obtained with the protocol reported above were analyzed in terms of intermolecular interactions with a program written in the macro language of CHARMM. The interface size (in $Å^2$) was calculated as the difference between the total accessible surface area of the separated partners without further relaxation and that of the complex, using the Lee and Richards algorithm,[93] with a probe radius of 1.4 Å. The gap volumes (in $Å^3$) of the different complexes were calculated, using the COOR VOL facility of CHARMM, as the difference between the total volume of the separated proteins without further refinement and that of the complex. The gap volume index is defined as the ratio of the gap volume (in $Å^3$) to the interface size (in $Å^2$).

The criteria used to assign the hydrogen bonds are those used in the HBPLUS program.[83,84] The distance between donor and acceptor must be lower than 3.9 Å and the distance between hydrogen and acceptor lower than 2.5 Å. The minimum angles are 90° for the donor−hydrogen-acceptor angle, the donor−acceptor−acceptor antecedent angle, and the hydrogen-acceptor−acceptor antecedent angle.

Residues were considered to be in the protonation state similar to that of free amino acids without p*K* shift at pH 7. Salt bridges were assigned when a group bearing a net charge of ±1 was located less than 3.0 Å from a group bearing a charge of opposite sign. Electrostatic clash was defined by the presence of two groups bearing a net charge of ±1 of identical sign located at a distance below 4.0 Å.

Protein Production and Characterization. (a) DLM-1 Domains: DNA Constructs and Protein Purification. Production of recombinant DLM-1 domains was adapted from previous reports.[59,94] A synthetic gene encoding mutDLM1 (Figure 3) cloned in pcDNA3.1 (Genotech) was recloned in the expression vector pET28a (Novagen) using the *Nde*I and *Hin*dIII restriction sites. The resulting construction encodes the fusion protein $(His)_6$-Leu-Val-Pro-Arg-Gly-Ser(thrombin site)-His-Met-mutDLM-1 (residues 8−70). Mutagenesis on this construct yielded a construct encoding the fusion protein $(His)_6$-LeuValProArgGlySer-(thrombin site)-HisMet-wtDLM-1 (residues 8−70). Starting from this initial construct, mutagenesis was done using the PCR technique (QuickChange Site-Directed Mutagenesis Kit, Stratagene) to obtain the different variants including the native form. Nucleotide sequences of all constructs were verified using the ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

For production of recombinant DLM-1 domains, 500 mL of LB medium containing 30 μg/mL kanamycin was seeded with 5 mL of an overnight culture of BL21-DE3 *Escherichia coli* strain transformed with pET28a-DLM-1 and incubated at 37 °C. When absorbance at 600 nm was equal to 0.8, expression of recombinant proteins was induced by adding 400 μM IPTG. Incubation was continued for 3 h at 37 °C, and cells were harvested by centrifugation. The pellet was resuspended in 20 mL of buffer A (50 mM Tris-HCl, pH 8, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 20 μg/mL Benzonase (Novagen)), and cells were disrupted using a French Press. The lysate was cleared by centrifugation (1 h at 16 000 rpm, JA20 rotor) and the supernatant incubated with 4 mL of Ni-NTA resin (Qiagen) for 1 h at 4 °C. The resin was washed in a column with 50 mL of buffer B (50 mM Tris-HCl, pH 8, 1 M NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol). $(His)_6$-tagged fusion proteins were then eluted with 20 mL of buffer C (50 mM Tris-HCl, pH 8, 300 mM NaCl, 200 mM imidazole, 5 mM β-mercaptoethanol), dialyzed overnight against buffer D (20 mM Tris-HCl, pH 8, 150 mM NaCl, 2 mM DTT), and cleaved with 1 unit of thrombin/100 μg of protein (Calbiochem) for 24 h at 22 °C. Then, 10 mL of 500 mM sodium phosphate buffer, pH 4.2, was added, and the solution was saturated under argon before addition of TCEP (1 equimolar). After 2 h of incubation at room temperature, DLM-1 domains were purified by RP-HPLC on a C18 column (Vydac, 22 × 250 mm) using a 92-min linear gradient of 20−100% solvent B (60% acetonitrile, 0.085% TFA) in solvent A (0.1% TFA) at a flow rate of 10 mL/min. Purified proteins were lyophilized and then solubilized in water for further studies.

(b) Chemical Synthesis. Proteins 1H5O, 1CKK, 1HZ6, and their mutants were obtained by solid-phase synthesis on an automated peptide synthetizer (model 433A, Applied Biosystems, Foster City, CA) by using the Fmoc/tBu strategy and a modified program. As a standard procedure, 0.1 mmol of each Fmoc-C-terminal preloaded amino-acid (see table in Supporting Information), a 10-fold excess of each Fmoc amino acid, DCCI/HOAt activation, and DIEA (0.1 mmol)/Triton X-100 (2% v/v)/LiCl (0.5 M)/NMP coupling conditions were used. After drying under vacuum, the mini-protein was liberated from resin and protecting groups by acidic cleavage with 10 mL of TFA/TIS/H_2O (95/2.5/2.5 v/v/v) for 2 h at room temperature. The resin was then filtered, and the free mini-protein was precipitated in 100 mL of ice-cold TBME. After centrifugation at 3000 rpm/15 min/4 °C, the solid was washed with precooled TBME (50 mL), again centrifugated, and finally dissolved in 20% acetic acid (20 mL) and freeze-dried. The crude proteins were submitted to RP-HPLC purification.

(c) Purification by RP-HPLC. The chemically synthetized mini-proteins were then purified to homogeneity by RP-HPLC on a Vydac C18 semipreparative column (particle diameter, 5 μm; pore size, 300 Å; column size, 10 × 250 mm) using nonlinear gradients (see table in Supporting Information) of eluent B (0.1% TFA, 90% acetonitrile in water) in eluent A (0.1% TFA, 10% acetonitrile in water) at a flow rate of 3 mL/min. Detection was achieved at 220 nm by a Merck L4000 UV detector equipped with a 5 mm cell. The homogeneity of the purified mini-proteins was assessed by analytical RP-HPLC (with the same gradient used for purification at a flow rate of 1 mL/min), mass spectrometry, and amino acid analysis. The physicochemical characteristics of each mini-protein are reported in the Supporting Information.

(d) Disulfide Pairing. For 1H5O and its mutants, the disulfide pairing was achieved at room temperature in degassed 50 mM Na_2HPO_4/NaH_2PO_4, pH 7.8, containing 5 mM reduced and 0.5 mM oxidized glutathione and protein at a final concentration of 1 mg/mL. Monitoring of disulfide pairing was achieved following the disappearance of the fully reduced protein by analytical RP-HPLC. Completion was observed after 3 h at room temperature. The solution was then acidified to pH 2 with TFA/H_2O (50/50 v/v) and purified as described in the "Purification by RP-HPLC" section. The homogeneity of the purified oxidized proteins was assessed by analytical RP-HPLC, mass spectrometry, and amino acid analysis. The physicochemical characteristics of each mini-protein are reported in the Supporting Information. For 1H5O (wild type and mutants), no disulfides assignments were achieved, but CD spectra between wild type and mutants were compared and no differences were observed.

(e) Mini-proteins' Concentrations. Concentrations were determined by recording UV spectra between 220−350 nm and measuring the absorbance at 280 mm on a Beckmann spectrometer at 25 °C, using a molar extinction coefficient calculated on the basis of the known amino acid content. Molar extinction coefficients, theoretical and experimental, for each protein are shown in the Supporting Information.

(f) CD Experiments. Far-UV CD spectra were recorded with a CDMax dichrograph (Jobin Yvon, Longjumeau, France), equipped with a thermostated cell holder and a PC operating with a CDMax data acquisition and manipulation program. Proteins were dissolved at a concentration ranging from 5 to 20 μM in 4 mM Tris buffer, pH 6.8. Spectra were run at 21 °C from 180 to 250 nm using a 0.1 cm quartz cell. Each spectrum represented the average of five spectra, obtained with an integration time of 1 s every 1 nm. Spectra were smoothed

(93) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(94) Schwartz, T.; Lowenhaupt, K.; Kim, Y-G.; Li, L.; Brown, B. A., II; Herbert, A.; Rich, A. *J. Biol. Chem.* **1999**, *274*, 2899−2906.

using the instrument software by taking a sliding average over nine data points. Spectra are reported in mean residue molar ellipticity. Additional spectra were recorded by adding 5−20% trifluoroethanol to the solution, or 10−100 mM potassium chloride. For the miniproteins 1H5O (wild-type and mutants), far-UV CD spectra were recorded with a JASCO J-815 instrument (Jasco France, Nantes, France) equipped with a Peltier cell holder and a PC operating with a Spectra Manager data acquisition and manipulation program. Spectra were run at 20 °C from 180 to 250 nm using a 0.1 cm quartz cell. The spectra were obtained using a 1 s integration time every 0.2 nm.

(g) Crystallization of 1J75_m(K130A) and Data Collection. Oligonucleotide 5′TCGCGCG (Genosphere Biotechnologies) was solubilized in 10 mM HEPES, pH 7.5/20 mM NaCl. Protein 1J75_m-(K130A) was solubilized in 5 mM HEPES, pH 7.5/0.5 M NaCl/5 mM DTT. 1J75_m(K130A) and DNA Z were mixed in equimolar amounts to concentrations of 1.15 mM each. Crystals were grown at 17 °C by seating-drop vapor diffusion experiments. Two crystal forms were obtained, needle- or prism-shaped, in the same conditions: 25% PEG4K, 0.2 M Na/K phosphate, 0.1 M MES, pH 6.0. Both crystals were harvested in a cryo-solution of 24% PEG4K/20% ethylene glycol and flash-frozen in liquid ethane. For each crystal form, a full data set was collected at the European Synchrotron Radiation Facility (ESRF, Grenoble), on beamline ID29, on a Quantum detector 315r at a wavelength of 0.976 Å. Data were integrated with the MOSFLM program as implemented in CCP4 (Collaborative Computational Project, no. 4, 1994). The diffraction power of each crystal form was different, leading to a maximum resolution of 1.7 Å for the prismatic crystals and 2.8 Å for the needles.

(h) Structural Determination and Refinement. The structure of the 1J75_m(K130A)-DNA Z complex was solved by molecular replacement using the MOLREP program from CCP4,[95] using the high-resolution data set and the DLM-1-DNA Z complex (PDB code 1J75[59]) as search model. After independent searches for the protein and DNA moieties of the complex, molecular replacement solution was compatible with two proteins and two DNA molecules per asymmetric unit. The solution clearly selected the P61 enantiomeric space group and reconstituted a complex similar to that previously observed.[59] The correctness of this solution was further assessed by the estimation of crystal solvent content in the P61 space group (MATTHEWS program from CCP4; Matthews, 1968) by a quick drop of both *R*-factor and *R*-free after an initial round of rigid-body refinement (REFMAC program from CCP4[96]) and by the observation of Watson−Crick base-paired DNA in the asymmetric unit. Refinement was further performed using REFMAC with successive introduction of water molecules and TLS parameters. The four molecules were initially restrained with non-crystallographic symmetry, and TLS parameters were applied using four rigid groups (one for each molecule). In order to validate the final structure, systematic omit maps were generated by removing part of the complex, REFMAC cycling, and weighted mapping (Supporting Information).

(i) Kv1.2 Potassium Channel Binding Assays. Potassium channel binding activities of proteins were tested by competition experiments (Racape et al., 2002) using as a tracer ^{125}I-BgK(W5Y/Y26F), an analogue of the sea anemone protein BgK that can be radiolabeled without loss of biological activity.[51] All these assays were carried out at room temperature in a medium consisting of 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM KCl, and 0.1% bovine serum albumin. Briefly, membrane preparations from transfected tsA cells producing Kv1.2 channels[51] (90 μL) were incubated with 50−100 pM ^{125}I-BgK-(W5Y/Y26F) in the absence or presence of increasing concentrations of the proteins to be tested. At the end of the incubation period, samples were filtered through Whatman GF/C glass fiber filters presoaked with 0.5% (w/v) polyethylenimine (Sigma), and filters were rinsed three times with 3 mL of ice-cold buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl). Data were analyzed as previously reported.[51]

(j) Coordinates. The atomic coordinates of the structure of 1J75_m_K130A-DNA Z have been deposited in the Protein Data Bank (accession code 2HEO).

Abbreviations. IPTG, isopropyl-β-D-thiogalactopyranoside; PMSF, phenylmethylsulfonyl fluoride; TCEP, tris(2-carboxyethyl)phosphine; Fmoc, fluorenylmethyloxycarbonyl; DCC, *N*,*N*′-dicyclohexylcarbodiimide; HOAt, 1-hydroxy-7-azabenzotriazole; DIEA, *N*,*N*-diisopropylethylamine; NMP, *N*-methylpyrrolidone; TFA, trifluoroacetic acid; TIS, triisopropylsilane; TBME, *tert*-butyl methyl ether; RP-HPLC, reverse-phase high-performance liquid chromatography; CD, circular dichroism; UV, ultraviolet; TFE, trifluoroethanol; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MES, 2-(*N*-morpholino)ethanesulfonic acid; DTT, dithiothreitol; PEG, polyethylene glycol; DNA, deoxynucleic acid; tBu, *tert*-butyl.

Acknowledgment. We thank the staff members of beamline ID29 at the European Synchrotron Radiation Facility (ESRF) Grenoble for their help in collecting data. We thank B. Gilquin for providing us with the coordinates of the BgK/S5−S6 part of the Kv1.1 channel. Dr. R. Thai is acknowledged for the mass spectroscopy analysis. Dr L. Le Clainche is acknowledged for his help in recording the CD spectra.

Supporting Information Available: Complete refs 23 and 90; chemical synthesis, purification, and characterization of proteins reported, including CD spectra; weighted $2F_o - F_c$ omit maps calculated for 1J75_m_K130A. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0646491

(95) Vagin, A.; Teplyakov, A. *J. Appl. Crystallogr.* **1997**, *30*, 1022−1025.
(96) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D* **1997**, *53*, 240−255.